Escudero Escorza, T. (2003). Desde los tests hasta la investigación evaluativa actual. Un siglo, el XX, de inteso desarrollo de la evaluación en educación. *RELIEVE*:, *v. 9*, n. 1, p. 11-43.
http://www.uv.es/RELIEVE/v9n1/RELIEVEv9n1_1.htm

RELIEVE

Revista ELectrónica de Investigación y EValuación Educativa

DESDE LOS TESTS HASTA LA INVESTIGACIÓN EVALUATIVA ACTUAL. UN SIGLO, EL XX, DE INTENSO DESARROLLO DE LA EVALUACIÓN EN EDUCACIÓN

(From tests to current evaluative research. One century, the XXth, of intense development of evaluation in education)

por

Articles record
About authors
HTML format

Tomás Escudero Escorza
(tescuder@posta.unizar.es)

Ficha artículo
Sobre los autores
Formato HTML

Abstract

This article presents a review and state of art about development in educational evaluation in the XXth century. The main theoretical proposals are commented

Resumen

Este artículo presenta una revisión crítica del desarrollo histórico que ha tenido el ámbito de la evaluación educativa durante todo el siglo XX. Se analizan los principales propuestas teóricas planteadas.

Keywords

Evaluation, Evaluation Research, Evaluation Methods; Formative Evaluation Summative Evaluation, Testing, Program Evaluation

Descriptores

Evaluación, Investigación evaluativa, Métodos de Evaluación, Evaluación Formativa, Evaluación Sumativa, Test, Evaluación de Programas

Introducción

En cualquier disciplina, la aproximación histórica suele ser una vía fundamental para comprender su concepción, estatus, funciones, ámbito, etc. Este hecho es especialmente evidente en el caso de la evaluación, pues se trata de una disciplina que ha sufrido profundas transformaciones conceptuales y funcionales a lo largo de la historia y, sobre todo, a lo largo del siglo XX, en el que principalmente ubicamos nuestro análisis. En este sentido, la aproximación diacrónica al concepto resulta imprescindible.

El análisis la vamos a llevar a cabo basándonos en tres planteamientos que podríamos tachar de clásicos en la reciente literatura sobre el tema y que usamos indistintamente, aunque no tenemos la pretensión de ofrecer un planteamiento de síntesis, sino de utilización cabal de todos ellos, puesto que los tres planteamientos inciden sobre los mismos momentos y movimientos claves.

Un planteamiento, quizás el más utilizado en nuestro contexto (Mateo y otros, 1993; Hernández, 1993), es el que ofrecen Madaus, Scri-

ven, Stufflebeam y otros autores, que en sus trabajos suelen establecer seis épocas, empezando su análisis desde el siglo XIX (Stufflebeam y Shinkfield, 1987; Madaus y otros, 1991). Nos hablan de: a) época de la reforma (1800-1900), b) época de la eficiencia y del «testing» (1900-1930), c) época de Tyler (1930-1945), d) época de la inocencia (1946-1956), e) época de la expansión (1957-1972) y f) época de la profesionalización (desde 1973), que enlaza con la situación actual.

Otros autores como Cabrera (1986) y Salvador (1992) citan tres grandes épocas, tomando como punto de referencia central la figura de Tyler en el segundo cuarto del Siglo XX. A la época de Tyler se le denomina de nacimiento, a las anteriores de precedentes o antecedentes y a la posterior de desarrollo.

Guba y sus colaboradores, sobre todo Yvonna Lyncoln, destacan distintas generaciones. Ahora estaríamos en la cuarta (Guba y Lincoln, 1989), que según ellos se apoya en el enfoque paradigmático constructivista y en las necesidades de los «stakeholders» (demandantes e implicados en la evaluación), como base para determinar la información que se necesita. La primera generación es la de la medición, que llega hasta el primer tercio de este siglo, la segunda es la de la descripción y la tercera la del juicio o valoración.

Tras el análisis histórico, como complemento y como revisión de síntesis del mismo, ofrecemos un sucinto resumen de los enfoques evaluativos más relevantes, de los distintos modelos y planteamientos que, con mayor o menor fuerza, vienen a nuestra mente cuando intentamos acotar lo que es hoy en día la investigación evaluativa en educación

1. Precedentes: Antes de los «tests» y de la medición

Desde la antigüedad se han venido creando y usando procedimientos instructivos en los que los profesores utilizaban referentes implícitos, sin una teoría explícita de evaluación, para valorar y, sobre todo, diferenciar y seleccionar a estudiantes. Dubois (1970) y Coffman (1971) citan los procedimientos que se empleaban en la China imperial, hace más de tres mil años, para seleccionar a los altos funcionarios. Otros autores como Sundbery (1977) hablan de pasajes evaluadores en la Biblia, mientras Blanco (1994) se refiere a los exámenes de los profesores griegos y romanos. Pero según McReynold (1975), el tratado más importante de evaluación de la antigüedad es el Tetrabiblos, que se atribuye a Ptolomeo. También Cicerón y San Agustín introducen en sus escritos conceptos y planteamientos evaluadores.

En la Edad Media se introducen los exámenes en los medios universitarios con carácter más formal. Hay que recordar los famosos exámenes orales públicos en presencia de tribunal, aunque sólo llegaban a los mismos los que contaban con el visto bueno de sus profesores, con lo que la posibilidad de fracaso era prácticamente inexistente. En el Renacimiento se siguen utilizando procedimientos selectivos y Huarte de San Juan, en su Examen de ingenios para las ciencias, defiende la observación como procedimiento básico de la evaluación (Rodríguez y otros, 1995).

En el siglo XVIII, a medida que aumenta la demanda y el acceso a la educación, se acentúa la necesidad de comprobación de los méritos individuales y las instituciones educativas van elaborando e introduciendo normas sobre la utilización de exámenes escritos (Gil, 1992).

Entrado el siglo XIX se establecen los sistemas nacionales de educación y aparecen los diplomas de graduación, tras la superación de exámenes (exámenes del Estado). Según Max Weber (Barbier, 1993), surge un sistema de exámenes de comprobación de una preparación específica, para satisfacer las necesidades de una nueva sociedad jerárquica y burocratizada. En los Estados Unidos, en 1845, Horace Mann comienza a utilizar las primeras técnicas evaluativas del tipo «tests» escritos, que se extienden a las escuelas de Boston, y que inician el camino hacia referentes más objetivos y explícitos con relación a determinadas destrezas lecto-escritoras. Sin embargo, no se trata toda-

vía de una evaluación sustentada en un enfoque teórico, sino más bien, algo que responde a prácticas en buena medida rutinarias y con frecuencia basadas en instrumentos poco fiables.

Al final del siglo XIX, en 1897, aparece un trabajo de J. M. Rice, que se suele señalar como la primera investigación evaluativa en educación (Mateo y otros, 1993). Se trataba de un análisis comparativo en escuelas americanas sobre el valor de la instrucción en el estudio de la ortografía, utilizando como criterio las puntuaciones obtenidas en los tests.

2. Los tests psicométricos

En el contexto anterior, a finales del siglo XIX, se despierta un gran interés por la medición científica de las conductas humanas. Esto es algo que se enmarca en el movimiento renovador de la metodología de las ciencias humanas, al asumir el positivismo de las ciencias físico-naturales. En este sentido, la evaluación recibe las mismas influencias que otras disciplinas pedagógicas relacionadas con procesos de medición, como la pedagogía experimental y la diferencial (Cabrera, 1986).

La actividad evaluativa se verá condicionada de forma decisiva por diversos factores que confluyen en dicho momento, tales como:

a) El florecimiento de las *corrientes filosóficas positivistas y empíricas,* que apoyaban a la observación, la experimentación, los datos y los hechos como fuentes del conocimiento verdadero. Aparece la exigencia del rigor científico y de la objetividad en la medida de la conducta humana (Planchard, 1960) y se potencian las pruebas escritas como medio para combatir la subjetividad de los exámenes orales (Ahman y Cook, 1967).

b) La influencia de las teorías evolucionistas y los trabajos de Darwin, Galton y Cattel, apoyando la medición de las características de los individuos y las diferencias entre ellos.

c) El desarrollo de los métodos estadísticos que favorecía decisivamente la orientación métrica de la época (Nunnally, 1978).

d) El desarrollo de la sociedad industrial que potenciaba la necesidad de encontrar unos mecanismos de acreditación y selección de alumnos, según sus conocimientos.

Consecuentemente con este estado de cosas, en este periodo entre finales del siglo XIX y principios del XX, se desarrolla una actividad evaluativa intensa conocida como «testing», que se define por características como las siguientes:

- Medición y evaluación resultaban términos intercambiables. En la práctica sólo se hablaba de medición.
- El objetivo era detectar y establecer diferencias individuales, dentro del modelo del rasgo y atributo que caracterizaba las elaboraciones psicológicas de la época (Fernández Ballesteros, 1981), es decir, el hallazgo de puntuaciones diferenciales, para determinar la posición relativa del sujeto dentro de la norma grupal.
- Los tests de rendimiento, sinónimo de evaluación educativa, se elaboraban para establecer discriminaciones individuales, olvidándose en gran medida la representatividad y congruencia con los objetivos educativos. En palabras de Guba y Lincoln (1982), la evaluación y la medida tenían poca relación con los programas escolares. Los tests informaban algo sobre los alumnos, pero no de los programas con los que se les había formado.

En el campo educativo destacan algunos instrumentos de aquella época, como las escalas de escritura de Ayres y Freeman, de redacción de Hillegas, de ortografía de Buckingan, de cálculo de Wood, de lectura de Thorndike y McCall y de aritmética de Wood y McCall (Planchard, 1960; Ahman y Cook, 1967; Ebel, 1977).

Sin embargo, fue en los tests psicológicos donde los esfuerzos tuvieron mayor impacto, siendo probablemente la obra de Thorndike (1904) la de mayor influencia en los comienzos del siglo XX. En Francia destacan los trabajos de Alfred Binet, después revisados por Terman en la Universidad de Stanford, sobre tests de capacidades cognitivas. Ahora hablamos del Stanford-Binet, uno de los tests más conocidos en la historia de la psicometría.

Años más tarde, con las necesidades de reclutamiento en la Primera Guerra Mundial, Arthur Otis dirige un equipo que construye tests colectivos de inteligencia general (Alfa para lectoescritores y Beta para analfabetos) e inventarios de personalidad (Phillips, 1974).

Tras la contienda, los tests psicológicos se ponen al servicio de fines sociales. La década entre 1920 y 1930 marca el punto más alto del «testing», pues se idean multitud de tests estandarizados para medir toda clase de destrezas escolares con referentes objetivos externos y explícitos, basados en procedimientos de medida de la inteligencia, para utilizar con grandes colectivos de estudiantes.

Estas aplicaciones estandarizadas se acogen muy bien en los ámbitos educativos y McCall (1920) propone que los profesores construyan sus propias pruebas objetivas, para no tener que confiar exclusivamente en las propuestas por especialistas externos.

Este movimiento estuvo vigente en paralelo al proceso de perfeccionamiento de los tests psicológicos con el desarrollo de la estadística y del análisis factorial. El fervor por el «testing» decreció a partir de los años cuarenta e, incluso, empezaron a surgir algunos movimientos hipercríticos con estas prácticas.

Guba y Lincoln (1989) se refieren a esta evaluación como a la primera generación, que puede legítimamente ser denominada como la generación de la medida. El papel del evaluador era técnico, como proveedor de instrumentos de medición. Según estos autores, esta primera generación permanece todavía viva, pues todavía existen textos y publicaciones que utilizan de manera indisoluble evaluación y medida (Gronlund, 1985).

3. El nacimiento de la verdadera evaluación educativa: La gran reforma «tyleriana»

Antes de que llegara la revolución promovida por Ralph W. Tyler, en Francia se inicia en los años veinte una corriente independiente conocida como docimología (Pieron, 1968 y 1969; Bonboir, 1972), que supone un primer acercamiento a la verdadera evaluación educativa. Se criticaba, sobre todo, el divorcio entre lo enseñado y las metas de la instrucción. La evaluación se dejaba, en último término, en manos de una interpretación totalmente personal del profesor. Como solución se proponía: a) elaboración de taxonomías para formular objetivos, b) diversificación de fuentes de información, exámenes, expedientes académicos, técnicas de repesca y tests, c) unificación de criterios de corrección a partir del acuerdo entre los correctores de las pruebas y d) revisión de los juicios de valoración mediante procedimientos tales como la doble corrección, o la media de distintos correctores. Como puede verse, se trata de criterios en buena medida vigentes actualmente y, en algún caso, incluso avanzados.

Pero quien es tradicionalmente considerado como el padre de la evaluación educativa es Tyler (Joint Committee, 1981), por ser el primero en dar una visión metódica de la misma, superando desde el conductismo, muy en boga en el momento, la mera evaluación psicológica. Entre 1932 y 1940, en su famoso Eight-Year Study of Secondary Education para la Progressive Education Association, publicado dos años después (Smith y Tyler, 1942), plantea la necesidad de una evaluación científica que sirva para perfeccionar la calidad de la educación. La obra de síntesis la publica unos años después (Tyler, 1950), exponiendo de manera clara su idea de «curriculum», e integrando en él su método sistemático de evaluación educativa, como el proceso surgido para determinar en qué medida han sido alcanzados los objetivos

previamente establecidos (véase también Tyler, 1967 y 1969).

El «currículum» viene delimitado por las cuatro cuestiones siguientes:

a) ¿Qué *objetivos* se desean conseguir?

b) ¿Con qué *actividades* se pueden alcanzar?

c) ¿Cómo pueden *organizarse* eficazmente estas experiencias?

d) ¿Cómo se puede *comprobar* si se alcanzan los objetivos?

Y la buena evaluación precisa de las siguientes condiciones:

a) Propuesta clara de *objetivos.*

b) Determinación de las *situaciones* en las que se deben manifestar las conductas esperadas.

c) Elección de *instrumentos apropiados* de evaluación.

d) *Interpretación* de los resultados de las pruebas.

e) Determinación de la *fiabilidad* y *objetividad* de las medidas.

Esta evaluación ya no es una simple medición, porque supone un juicio de valor sobre la información recogida. Se alude, aunque sin desarrollar, a la toma de decisiones sobre los aciertos o fracasos de la programación, en función de los resultados de los alumnos, algo que retomarán otros importantes evaluadores como Cronbach y Sufflebeam unos años después.

Para Tyler, la referencia central en la evaluación son los objetivos preestablecidos, que deben ser cuidadosamente definidos en términos de conducta (Mager, 1962), teniendo en cuenta que deben marcar el desarrollo individual del alumno, pero dentro de un proceso socializador.

El objeto del proceso evaluativo es determinar el cambio ocurrido en los alumnos, pero su función es más amplia que el hacer explícito este cambio a los propios alumnos, padres y profesores; es también un medio para informar sobre la eficacia del programa educacional y también de educación continua del profesor. Se trata, según Guba y Lincoln (1989), de la segunda generación de la evaluación. Desgraciadamente, esta visión evaluativa global no fue suficientemente apreciada, ni explotada, por aquellos que utilizaron sus trabajos (Bloom y otros, 1975; Guba y Lincoln, 1982).

A pesar de lo anterior y de que las reformas tylerianas no siempre se aplicaron de inmediato, las ideas de Tyler fueron muy bien acogidas por los especialistas en desarrollo curricular y por los profesores. Su esquema era racional y se apoyaba en una tecnología clara, fácil de entender y aplicar (Guba y Lincoln, 1982; House, 1989) y encajaba perfectamente en la racionalidad del análisis de la tarea que comenzaba a usarse con éxito en ámbitos educativos militares (Gagné, 1971). En España, los planteamientos de Tyler se extendieron con la Ley General de Educación de 1970.

Tras la Segunda Guerra Mundial se produce un periodo de expansión y optimismo que Stufflebeam y Shinkfield (1987) no han dudado en calificar de «irresponsabilidad social», por el gran despilfarro consumista tras una época de recesión. Se trata de la etapa conocida como la de la inocencia (Madaus y otros, 1991). Se extienden mucho las instituciones y servicios educativos de todo tipo, se producen cantidad de tests estandarizados, se avanza en la tecnología de la medición y en los principios estadísticos del diseño experimental (Gulliksen, 1950; Lindquist, 1953; Walberg y Haertel, 1990) y aparecen las famosas taxonomías de los objetivos educativos (Bloom y otros, 1956; Krathwohl y otros, 1964). Sin embargo, en esta época, la aportación de la evaluación a la mejora de la enseñanza es escasa debido a la carencia de planes coherentes de acción. Se escribe mucho de evaluación, pero con escasa influencia en el perfeccionamiento de la labor instruccional. El verdadero desarrollo de las propuestas

tylerianas vino después (Taba, 1962; Popham y Baker, 1970; Fernández de Castro, 1973).

Ralph W. Tyler murió el 18 de febrero de 1994, superados los noventa años de vida, tras siete décadas de fructíferas aportaciones y servicios a la evaluación, a la investigación y a la educación en general. Unos meses antes, en abril de 1993, Pamela Perfumo, una estudiante graduada de la Universidad de Stanford, entrevistó a Tyler con el propósito de conocer su pensamiento acerca del actual desarrollo de la evaluación y de los temas controvertidos alrededor de la misma. Esta entrevista, convenientemente preparada, fue presentada el 16 de abril de 1993 en la Conferencia de la AERA que tuvo lugar en Atlanta. Horowitz (1995) analiza el contenido y el significado de la citada entrevista, destacando, entre otros, los siguientes aspectos en el pensamiento de Tyler al final de sus días:

a) Necesidad de analizar cuidadosamente los propósitos de la evaluación, antes de ponerse a evaluar. Los actuales planteamientos de evaluaciones múltiples y alternativas deben ajustarse a este principio

b) El propósito más importante en la evaluación de los alumnos es guiar su aprendizaje, esto es, ayudarles a que aprendan. Para ello es necesaria una evaluación comprensiva de todos los aspectos significativos de su rendimiento; no basta con asegurarse que hacen regularmente el trabajo diario.

c) El «portfolio» es un instrumento valioso de evaluación, pero depende de su contenido. En todo caso, hay que ser cauteloso ante la preponderancia de un solo procedimiento de evaluación, incluyendo el «portfolio», por su incapacidad de abarcar todo el espectro de aspectos evaluables.

d) La verdadera evaluación debe ser idiosincrásica, adecuada a las peculiaridades del alumno y el centro. En rigor, la comparación de centros no es posible.

e) Los profesores deben rendir cuentas de su acción educativa ante los padres de los alumnos. Para ello, es necesario interaccionar con ellos de manera más frecuente y más informal.

Medio siglo después de que Tyler revolucionara el mundo de la evaluación educativa, se observa la fortaleza, coherencia y vigencia de su pensamiento. Como acabamos de ver, sus ideas básicas, convenientemente actualizadas, se entroncan fácilmente en las corrientes más actuales de la evaluación educativa.

4. El desarrollo de los sesenta

Los años sesenta traerán nuevos aires a la evaluación educativa, entre otras cosas porque se empezó a prestar interés por algunas de las llamadas de atención de Tyler, relacionadas con la eficacia de los programas y el valor intrínseco de la evaluación para la mejora de la educación.

En esa época surge un cierto conflicto entre la sociedad americana y su sistema educativo, sobre todo porque los rusos iban por delante en la carrera especial, tras el lanzamiento del Sputnik por la URSS en 1957. Aparece un cierto desencanto con la escuela pública y crece la presión por la rendición de cuentas (MacDonald, 1976; Stenhouse, 1984). En 1958 se promulga una nueva ley de defensa educativa que proporciona muchos programas y medios para evaluarlos. En 1964 se establece el Acta de educación primaria y secundaria (ESEA) y se crea el National Study Comitte on Evaluation, creándose una nueva evaluación no sólo de alumnos, sino orientada a incidir en los programas y en la práctica educativa global (Mateo y otros, 1993; Rodríguez y otros, 1995).

Para mejorar la situación y retomar la hegemonía científica y educativa, fueron muchos los millones de dólares que desde los fondos públicos se destinaron a subvencionar nuevos programas educativos e iniciativas del personal de las escuelas públicas americanas encaminadas a mejorar la calidad de la enseñanza. (Popham, 1983; Rutman y Mowbray, 1983; Weiss, 1983). Este movimiento se vio también potenciado por el desarrollo de nuevos medios tecnológicos (audiovisuales, ordenadores...) y

nológicos (audiovisuales, ordenadores...) y el de la enseñanza programada, cuyas posibilidades educativas despertaron el interés entre los profesionales de la educación (Rosenthal, 1976).

De la misma forma que la proliferación de programas sociales en la década anterior había impulsado la evaluación de programas en el área social, los años sesenta serán fructíferos en demanda de evaluación en el ámbito de la educación. Esta nueva dinámica en la que entra la evaluación, hace que, aunque ésta se centraba en los alumnos como sujeto que aprende, y el objeto de valoración era el rendimiento de los mismos, sus funciones, su enfoque y su última interpretación variará según el tipo de decisión buscada.

Buena parte de culpa de este fuerte ímpetu evaluador americano se debió a la ya citada aprobación de la «Elementary and Secondary Act» (ESEA) en 1965 (Berk, 1981; Rutman, 1984). Con esta ley se puso en marcha el primer programa significativo para la organización de la educación en el ámbito federal de los Estados Unidos, y se estipuló que cada uno de los proyectos realizados con el apoyo económico federal debía ser anualmente evaluado, a fin de justificar subvenciones futuras.

Junto al desencanto de la escuela pública, cabe señalar la recesión económica que caracteriza los finales años sesenta, y, sobre todo, la década de los setenta. Ello hizo que la población civil, como contribuyentes, y los propios legisladores se preocupasen por la eficacia y el rendimiento del dinero que se empleaba en la mejora del sistema escolar. A finales de los años sesenta, y como consecuencia de lo anterior, entra en escena un nuevo movimiento, la era de la «Accountability», de la rendición de cuentas (Popham, 1980 y 1983; Rutman y Mowbray, 1983), que se asocia fundamentalmente a la responsabilidad del personal docente en el logro de objetivos educativos establecidos. De hecho, en el año 1973, la legislación de muchos estados americanos instituyó la obligación de controlar el logro de los objetivos educativos y la adopción de medidas correctivas en caso negativo (MacDonald, 1976; Wilson y otros, 1978). Es comprensible que, planteado así, este movimiento de rendición de cuentas, de responsabilidad escolar, diera lugar a una oleada de protestas por parte del personal docente.

Otra dimensión de la responsabilidad escolar nos la ofrece Popham (1980), cuando se refiere al movimiento de descentralización escolar durante los últimos años sesenta y principios de los setenta. Los grandes distritos escolares se dividieron en áreas geográficas más pequeñas, y, por consiguiente, con un control ciudadano más directo sobre lo que ocurría en las escuelas.

Como consecuencia de estos focos de influencia, se amplió considerablemente el fenómeno de la evaluación educativa. El sujeto directo de la evaluación siguió siendo el alumno, pero también todos aquellos factores que confluyen en el proceso educativo (el programa educativo en un sentido amplio, profesor, medios, contenidos, experiencias de aprendizaje, organización, etc.), así como el propio producto educativo.

Como resultado de estas nuevas necesidades de la evaluación, se inicia durante esta época un periodo de reflexión y de ensayos teóricos con ánimo de clarificar la multidimensionalidad del proceso evaluativo. Estas reflexiones teóricas enriquecerán decisivamente el ámbito conceptual y metodológico de la evaluación, lo que unido a la tremenda expansión de la evaluación de programas ocurrida durante estos años, dará lugar al nacimiento de esa nueva modalidad de investigación aplicada que hoy denominamos como investigación evaluativa.

Como hitos de la época hay que destacar dos ensayos por su decisiva influencia: el artículo de Cronbach (1963), Course improvement through evaluation, y el de Scriven (1967), The methodology of evaluation. La riqueza de ideas evaluativas expuestas en estos trabajos nos obligan a que, aunque brevemente, nos refiramos a ellas.

Del análisis que Cronbach del concepto, funciones y metodología de la evaluación, entresacamos las sugerencias siguientes:

a) *Asociar el concepto de evaluación a la toma de decisiones.* Distingue el autor tres tipos de decisiones educativas a las cuales la evaluación sirve: a) sobre el perfeccionamiento del programa y de la instrucción, b) sobre los alumnos (necesidades y méritos finales) y c) acerca de la regulación administrativa sobre la calidad del sistema, profesores, organización, etc. De esta forma, Cronbach abre el campo conceptual y funcional de la evaluación educativa mucho más allá del marco conceptual dado por Tyler, aunque en su línea de sugerencias.

b) La evaluación que se usa para *mejorar un programa mientras éste se está aplicando,* contribuye más al desarrollo de la educación que la evaluación usada para estimar el valor del producto de un programa ya concluido.

c) Poner en cuestión la necesidad de que los estudios evaluativos sean de tipo comparativo. Entre las objeciones a este tipo de estudios, el autor destaca el hecho de que, con frecuencia, las diferencias entre las puntuaciones promedio entre-grupos son menores que las intragrupos, así como otras referentes a las dificultades técnicas que en el marco educativo presentan los diseños comparativos. Cronbach aboga por unos criterios de comparación de tipo absoluto, reclamando la necesidad de una evaluación con referencia al criterio, al defender la valoración con relación a unos objetivos bien definidos y no la comparación con otros grupos.

d) Se ponen en cuestión los estudios a gran escala, puesto que las diferencias entre los tratamientos pueden ser muy grandes e impedir discernir con claridad las causas de los resultados. Se defienden los estudios más analíticos, bien controlados, que pueden usarse para comparar versiones alternativas de un programa.

e) Metodológicamente Cronbach propone que la evaluación debe incluir: 1) estudios de proceso –hechos que tienen lugar en el aula–; 2) medidas de rendimiento y actitudes –cambios observados en los alumnos– y 3) estudios de seguimientos, esto es, el camino posterior seguido por los estudiantes que han participado en el programa.

f) Desde esta óptica, las técnicas de evaluación no pueden limitarse a los tests de rendimiento. Los cuestionarios, las entrevistas, la observación sistemática y no sistemática, las pruebas de ensayo, según el autor, ocupan un lugar importante en la evaluación, en contraste al casi exclusivo uso que se hacía de los tests como técnicas de recogida de información.

Si estas reflexiones de Cronbach fueron impactantes, no lo fueron menos las del ensayo de Scriven (1967). Sus fecundas distinciones terminológicas ampliaron enormemente el campo semántico de la evaluación, a la vez que clarificaron el quehacer evaluativo. Destacamos a continuación las aportaciones más significativas:

a) Se establece de forma tajante la diferencia entre la evaluación como actividad metodológica, lo que el autor llama meta de la evaluación, y las funciones de la evaluación en un contexto particular. Así, la evaluación como actividad metodológica es esencialmente igual, sea lo que fuera lo que estemos evaluando. El objetivo de la evaluación es invariante, supone en definitiva el proceso por el cual estimamos el valor de algo que se evalúa, mientras que las funciones de la evaluación pueden ser enormemente variadas. Estas funciones se relacionan con el uso que se hace de la información recogida.

b) Scriven señala dos funciones distintas que puede adoptar la evaluación: la formativa y la sumativa. Propone el término de *evaluación formativa* para calificar aquel proceso de evaluación al servicio de un programa en desarrollo, con objeto de mejorarlo, y el término de *evaluación sumativa* para aquel proceso orientado a comprobar la eficacia del programa y tomar decisiones sobre su continuidad.

c) Otra importante contribución de Scriven es la crítica al énfasis que la evaluación da a la consecución de objetivos previamente establecidos, porque si los objetivos carecen de valor, no tiene ningún interés saber hasta qué punto se han conseguido. Resalta la necesidad de que la evaluación debe incluir tanto la evaluación de los propios objetivos como el determinar el grado en que éstos han sido alcanzados (Scriven, 1973 y 1974).

d) Clarificadora es también la distinción que hace Scriven entre evaluación intrínseca y evaluación extrínseca, como dos formas diferentes de valorar un elemento de la enseñanza. En una evaluación intrínseca se valora el elemento por sí mismo, mientras que en la evaluación extrínseca se valora el elemento por los efectos que produce en los alumnos. Esta distinción resulta muy importante a la hora de considerar el criterio a utilizar, pues en la evaluación intrínseca el criterio no se formula en términos de objetivos operativos, mientras que sí se hace en la evaluación extrínseca .

e) Scriven adopta una posición contraria a Cronbach, defendiendo el carácter comparativo que deben presentar los estudios de evaluación. Admite con Cronbach los problemas técnicos que los estudios comparativos entrañan y la dificultad de explicar las diferencias entre programas, pero Scriven considera que la evaluación como opuesta a la mera descripción implica emitir un juicio sobre la superioridad o inferioridad de lo que se evalúa con respecto a sus competidores o alternativas.

Estas dos aportaciones comentadas influyeron decisivamente en la comunidad de evaluadores, incidiendo no sólo en estudios en la línea de la investigación evaluativa, a la que se referían preferentemente, sino también en la evaluación orientada al sujeto, en la línea de evaluación como «assessment» (Mateo, 1986). Estamos ante la *tercera generación* de la evaluación que, según Guba y Lincoln (1989), se caracteriza por introducir la valoración, *el juicio,* como un contenido intrínseco en la evaluación. Ahora el evaluador no sólo analiza y describe la realidad, además, la valora, la juzga con relación a distintos criterios.

Durante estos años sesenta aparecen muchas otras aportaciones que va perfilando una nueva concepción evaluativa, que terminará de desarrollarse y, sobre todo, de extenderse en las décadas posteriores. Se percibe que el núcleo conceptual de la evaluación lo constituye la valoración del cambio ocurrido en el alumno como efecto de una situación educativa sistemática, siendo unos objetivos bien formulados el mejor criterio para valorar este cambio. Así mismo, se comienza a prestar atención no sólo a los resultados pretendidos, sino también a los efectos laterales o no pretendidos, e incluso a resultados o efectos a largo plazo (Cronbach, 1963; Glaser, 1963; Scriven, 1967; Stake, 1967).

A pesar de las voces críticas con la operativización de objetivos (Eisner, 1967 y 1969; Atkin, 1968), no sólo por la estructura de valor que en ello subyace, sino también por centrar la valoración del aprendizaje en los productos más fácilmente mensurables, a veces los más bajos en las taxonomías del dominio cognoscitivo, y de que se prestaba escasa atención a los objetivos del dominio afectivo, que presentan mayor dificultad de tratamiento operativo, el modelo evaluativo de Tyler se enriquecería mucho en estos años, con trabajos sobre los objetivos educativos que continuarían y perfeccionarían el camino emprendido en 1956 por Bloom y colaboradores (Mager, 1962 y 1973; Lindvall, 1964; Krathwohl y otros, 1964; Glaser, 1965; Popham, 1970; Bloom y otros, 1971; Gagné 1971). Entre otras cosas aparecieron nuevas ideas sobre la evaluación de la interacción en el aula y sobre sus efectos en los logros de los alumnos (Baker, 1969).

Stake (1967) propuso su modelo de evaluación, *The countenance model*, que sigue la línea de Tyler, pero es más completo al considerar las discrepancias entre lo observado y lo esperado en los «antecedentes» y «transacciones», y posibilitar algunas bases para elaborar

hipótesis acerca de las causas y los fallos en los resultados finales. En sus sucesivas propuestas, Stake se irá distanciando de sus posiciones iniciales.

Metfessell y Michael (1967) presentaron también un modelo de evaluación de la efectividad de un programa educativo en el cual, aún siguiendo el modelo básico de Tyler, proponían la utilización de una lista comprensiva de criterios diversos que los evaluadores podrían tener en cuenta en el momento de la valoración y, por consiguiente, no centrarse meramente en los conocimientos intelectuales alcanzados por los alumnos.

Suchman (1967) profundiza en la convicción de que la evaluación debe basarse en datos objetivos que sean analizados con metodología científica, matizando que la investigación científica es preferentemente teórica y, en cambio, la investigación evaluativa es siempre aplicada. Su principal propósito es descubrir la efectividad, éxito o fracaso de un programa al compararlo con los objetivos propuestos y, así, trazar las líneas de su posible redefinición. Esta investigación evaluativa para Suchman debe tener en cuenta: a) la naturaleza del destinatario del objetivo y la del propio objetivo, b) el tiempo necesario para que se realice el cambio propuesto, c) el conocimiento de si los resultados esperados son dispersos o concentrados y d) los métodos que han de emplearse para alcanzar los objetivos. Suchman, además, defiende la utilización de evaluadores externos para evitar todo tipo de tergiversación de los profesores muy implicados en los procesos instruccionales.

El énfasis en los objetivos y su medida traerá también la necesidad de una nueva orientación a la evaluación, la denominada *evaluación de referencia criterial*. La distinción introducida por Glaser (1963) entre mediciones referidas a normas y criterios tendrá eco al final de la década de los sesenta, precisamente como resultado de las nuevas exigencias que a la evaluación educativa se le planteaban. Así, por ejemplo, cuando Hambleton (1985) estudia las diferencias entre tests referidos al criterio y tests referidos a la norma, señala para los primeros, además de los conocidos objetivos de describir la ejecución del sujeto y tomar decisiones sobre si domina o no domina un contenido, otro objetivo como es el de valorar la eficacia de un programa.

Desde finales de los sesenta los especialistas se pronunciarán decisivamente a favor de la *evaluación criterial,* en cuanto que es el tipo de evaluación que suministra una información real y descriptiva del estatus del sujeto o sujetos respecto a los objetivos de enseñanza previstos, así como la valoración de ese estatus por comparación con un estándar o criterio de realizaciones deseables, siendo irrelevantes, al efecto de contraste, los resultados obtenidos por otros sujetos o grupo de sujetos (Popham, 1970 y 1983; Mager, 1973; Carreño, 1977; Gronlund, 1985).

En las prácticas evaluativas de esta década de los sesenta se observan dos niveles de actuación. Un nivel podemos calificarlo como la *evaluación orientada hacia los individuos,* fundamentalmente alumnos y profesores. El otro nivel, es el de la *evaluación orientada a la toma de decisiones sobre el «instrumento» o «tratamiento» o «programa» educativo.* Este último nivel, impulsado también por la evaluación de programas en el ámbito social, será la base para la consolidación en el terreno educativo de la evaluación de programas y de la investigación evaluativa.

5. Desde los años setenta: La consolidación de la investigación evaluativa

Si con algo se podría caracterizar las aportaciones teóricas que nos ofrecen los especialistas durante *los años setenta* es con la proliferación de toda clase de modelos evaluativos que inundan el mercado bibliográfico, *modelos de evaluación* que expresan la propia óptica del autor que los propone sobre *qué es* y *cómo* debe conducirse un proceso evaluativo. Se trata, por tanto, de una época caracterizada por la pluralidad conceptual y metodológica. Guba y Lincoln (1982) nos hablan de más de cuarenta modelos propuestos en estos años, y Mateo

(1986) se refiere a la *eclosión de modelos.* Estos enriquecerán considerablemente el vocabulario evaluativo, sin embargo, compartimos la idea de Popham (1980) de que algunos son demasiado complicados y otros utilizan una jerga bastante confusa.

Algunos autores como Guba y Lincoln (1982), Pérez (1983) y en alguna medida House (1989), tienden a clasificar estos modelos en dos grandes grupos, cuantitativos y cualitativos, pero nosotros pensamos con Nevo (1983) y Cabrera (1986) que la situación es mucho más rica en matices.

Es cierto que esas *dos tendencias* se observan hoy en las propuestas evaluativas, y que algunos modelos pueden ser representativos de ellas, pero los diferentes modelos, considerados particularmente, se diferencian más por destacar o enfatizar alguno o algunos de los componentes del proceso evaluativo y por la particular interpretación que a este proceso le dan. Es desde esta perspectiva, a nuestro entender, como los diferentes modelos deben ser vistos, y valorar así sus respectivas aportaciones en los terrenos conceptual y metodológico (Worthen y Sanders, 1973; Stufflebeam y Shinkfield, 1987; Arnal y otros, 1992; Scriven, 1994).

También son varios los autores (Lewy, 1976; Popham, 1980; Cronbach, 1982; Anderson y Ball, 1983; De la Orden, 1985) los que consideran los modelos no como excluyentes, sino más bien como complementarios y que el estudio de los mismos (al menos aquellos que han resultado ser más prácticos) llevará al evaluador a adoptar una visión más amplia y comprensiva de su trabajo. Nosotros, en algún momento nos hemos atrevido a hablar de enfoques modélicos, más que de modelos, puesto que es cada evaluador el que termina construyendo su propio modelo en cada investigación evaluativa, en función del tipo de trabajo y las circunstancias (Escudero, 1993).

En este movimiento de propuestas de modelos de evaluación cabe distinguir dos épocas con marcadas diferencias conceptuales y metodológicas. En una *primera época*, las propuestas seguían la línea expuesta por Tyler en su planteamiento, que ha venido a llamarse de *"Consecución de Metas".* Además de los ya citados de Stake y Metfessell y Michael, que corresponden a los últimos años sesenta, en esta época destacan la propuesta de Hammond (1983) y el *Modelo de Discrepancia* de Provus (1971). Para estos autores los objetivos propuestos siguen siendo el criterio fundamental de valoración, pero enfatizan la necesidad de aportar datos sobre la congruencia o discrepancia entre las pautas de instrucción diseñadas y la ejecución de las mismas en la realidad del aula.

Otros modelos consideran el proceso de evaluación al servicio de las instancias que deben tomar decisiones. Ejemplos notables de ellos son: probablemente el más famoso y utilizado de todos, el C.I.P.P. (contexto, input, proceso y producto), propuesto por Stufflebeam y colaboradores (1971) y el C.E.S. (toma sus siglas del Centro de la Universidad de California para el Estudio de la Evaluación) dirigido por Alkin (1969). La aportación conceptual y metodológica de estos modelos es valorada positivamente entre la comunidad de evaluadores (Popham, 1980; Guba y Lincoln, 1982; House, 1989). Estos autores van más allá de la evaluación centrada en resultados finales, puesto que en sus propuestas suponen diferentes tipos de evaluación, según las necesidades de las decisiones a las que sirven.

Una *segunda época* en la proliferación de modelos es la representada por los *modelos alternativos*, que con diferentes concepciones de la evaluación y de la metodología a seguir comienzan a aparecer en la segunda mitad de esta década de los setenta. Entre ellos destacan la *Evaluación Responsable* de Stake (1975 y 1976), a la que se adhieren Guba y Lincoln (1982), la *Evaluación Democrática* de MacDonald (1976), la *Evaluación Iluminativa* de Parlett y Hamilton (1977) y la *Evaluación como crítica artística* de Eisner (1985).

En líneas generales, este segundo grupo de modelos evaluativos enfatiza el papel de la *audiencia* de la evaluación y de la relación del evaluador con ella. La audiencia prioritaria de la evaluación en estos modelos no es quien

debe tomar las decisiones, como en los modelos orientados a la toma de decisiones, ni el responsable de elaborar los currículos u objetivos, como en los modelos de consecución de metas. La audiencia prioritaria son los propios participantes del programa. La relación entre el evaluador y la audiencia en palabras de Guba y Lincoln (1982) debe ser «transaccional y fenomenológica». Se trata de modelos que propugnan una evaluación de tipo etnográfica, de aquí que la metodología que consideran más adecuada es la propia de la antropología social (Parlett y Hamilton, 1977; Guba y Lincoln, 1982; Pérez 1983).

Este resumen de modelos de la época de eclosión es suficiente para aproximarnos al amplio abanico conceptual teórico y metodológico que hoy se relaciona con la evaluación. Ello explica que cuando Nevo (1983 y 1989) pretende realizar una conceptualización de la evaluación, a partir de la revisión de la literatura especializada, atendiendo a los tópicos ¿qué es la evaluación? ¿qué funciones tiene? ¿cuál es el objeto de evaluación?... no encuentra una única respuesta a estas cuestiones. Es fácilmente comprensible que las exigencias que plantea la evaluación de programas de una parte, y la evaluación para la toma de decisiones sobre los individuos de otra, conducen a una gran variedad de esquemas evaluativos reales utilizados por profesores, directores, inspectores y administradores públicos. Pero también es cierto que bajo esta diversidad subyacen diferentes concepciones teóricas y metodológicas sobre la evaluación. Diferentes concepciones que han dado lugar a una apertura y pluralidad conceptual en el ámbito de la evaluación en varios sentidos (Cabrera, 1986). A continuación destacamos los puntos mas sobresalientes de esta pluralidad conceptual.

a) *Diferentes conceptos de evaluación.* Por una parte existe la clásica definición dada por Tyler: *la evaluación como el proceso de determinar el grado de congruencia entre las realizaciones y los objetivos previamente establecidos,* a la que corresponden los modelos orientados hacia la consecución de metas. Contrasta esta definición con aquella más amplia que se propugna desde los modelos orientados a la toma de decisiones: *la evaluación como el proceso de determinar, obtener y proporcionar información relevante para juzgar decisiones alternativas*, defendida por Alkin (1969), Stufflebeam y otros (1971), MacDonald (1976) y Cronbach (1982).

Además, el concepto de evaluación de Scriven (1967), como el proceso de estimar el valor o el mérito de algo, es retomado por Cronbach (1982), Guba y Lincoln (1982), y House (1989), con objeto de señalar las diferencias que comportarían los juicios valorativos en caso de estimar el *mérito* (se vincularía a características intrínsecas de lo que se evalúa) o el *valor* (se vincularía al uso y aplicación que tendría para un contexto determinado).

b) *Diferentes criterios.* De las definiciones apuntadas anteriormente se desprende que el criterio a utilizar para la valoración de la información también cambia. Desde la óptica de la consecución de metas, una buena y operativa definición de los objetivos constituye el criterio fundamental. Desde la perspectiva de las decisiones y situados dentro de un contexto político, Stufflebeam y colaboradores, Alkin y MacDonald llegan a sugerir incluso la no valoración de la información por parte del evaluador, siendo el que toma las decisiones el responsable de su valoración.

Las definiciones de evaluación que acentúan la determinación del «mérito» como objetivo de la evaluación, utilizan criterios estándares sobre los que los expertos o profesionales están de acuerdo. Se trata de modelos relacionados con la acreditación y el enjuiciamiento profesional (Popham, 1980).

Los autores (Stake, 1975; Parlett y Hamilton, 1977; Guba y Lincoln, 1982; House, 1983) que acentúan el proceso de evaluación al servicio de determinar el «valor» más que el «mérito» de la entidad u objeto evaluado, abogan por que el criterio de valoración fundamental sean las necesidades contextuales en las que ésta se inserta. Así, Guba y Lin-

coln (1982) refieren los términos de la comparación valorativa; de un lado, las características del objeto evaluado y, de otro, las necesidades, expectativas y valores del grupo a los que les afecta o con los que se relaciona el objeto evaluado.

c) *Pluralidad de procesos evaluativos* dependiendo de la percepción teórica que sobre la evaluación se mantenga. Los modelos de evaluación citados y otros más que pueden encontrarse en la bibliografía, representan diferentes propuestas para conducir una evaluación.

d) *Pluralidad de objetos de evaluación.* Como dice Nevo (1983 y 1989), existen dos conclusiones importantes que se obtienen de la revisión de la bibliografía sobre la evaluación. Por un lado, cualquier cosa puede ser objeto de evaluación y ésta no debería limitarse a estudiantes y profesores y, por otro, una clara identificación del objeto de evaluación es una importante parte en cualquier diseño de evaluación.

e) *Apertura,* reconocida en general por todos los autores, de la información necesaria en un proceso evaluativo para dar cabida no sólo a los resultados pretendidos, sino a los *efectos posibles* de un programa educativo, sea pretendido o no. Incluso Scriven (1973 y 1974) propone una evaluación en la que no se tenga en cuenta los objetivos pretendidos, sino valorar todos los efectos posibles. Apertura también respecto a la recogida de información no sólo del producto final, sino también sobre el *proceso* educativo. Y apertura en la consideración de diferentes resultados de *corto* y *largo alcance.* Por último, apertura también en considerar no sólo resultados de tipo cognitivo, sino también afectivos (Anderson y Ball, 1983).

f) Pluralidad también reconocida de las *funciones* de la evaluación en el ámbito educativo, recogiéndose la propuesta de Scriven entre evaluación formativa y sumativa, y añadiéndose otras de tipo socio-político y administrativas (Nevo, 1983).

g) Diferencias en el papel jugado por *el evaluador,* lo que ha venido a llamarse *evaluación interna* vs. *evaluación externa.* No obstante, una relación directa entre el evaluador y las diferentes audiencias de la evaluación es reconocida por la mayoría de los autores (Nevo, 1983; Weiss, 1983; Rutman, 1984).

h) *Pluralidad de audiencia* de la evaluación y, por consiguiente, *pluralidad en los informes* de evaluación. Desde informes narrativos, informales, hasta informes muy estructurados (Anderson y Ball, 1983).

i) *Pluralidad metodológica.* Las cuestiones metodológicas surgen desde la dimensión de la evaluación como investigación evaluativa, que viene definida en gran medida por la diversidad metodológica.

El anterior resumen recoge las aportaciones a la evaluación en los años setenta y ochenta, la época que se ha denominado *época de la profesionalización* (Stufflebeam y Skinkfield, 1987; Madaus y otros, 1991; Hernández, 1993; Mateo y otros, 1993), en la que además de los innumerables modelos de los setenta, se profundizó en los planteamientos teóricos y prácticos y se consolidó la evaluación como *investigación evaluativa* en los términos antes definida. En este contexto, lógicamente, aparecen muchas nuevas revistas especializadas como *Educational Evaluation and Policy Analysis*, *Studies in Evaluation, Evaluation Review*, *New Directions for Program Evaluation*, *Evaluation and Program Planning*, *Evaluation News*,..., se fundan asociaciones científicas relacionadas con el desarrollo de la evaluación y las universidades empiezan a ofrecer cursos y programas de investigación evaluativa, no sólo en postgrados y programas de doctorado, sino también en planes de estudio para titulaciones de primer y segundo ciclos.

6. La cuarta generación según Guba y Lincoln

A finales de los ochenta, tras todo este desarrollo antes descrito, Guba y Lincoln (1989) ofrecen una alternativa evaluadora, que

ofrecen una alternativa evaluadora, que denominan *cuarta generación*, pretendiendo superar lo que según estos autores son deficiencias de las tres generaciones anteriores, tales como una visión gestora de la evaluación, una escasa atención al pluralismo de valores y un excesivo apego al paradigma positivista. La alternativa de Guba y Lincoln la denominan *respondente y constructivista*, integrando de alguna manera el enfoque respondente propuesto en primer lugar por Stake (1975), y la epistemología *postmoderna* del constructivismo (Russell y Willinsky, 1997). Las demandas, las preocupaciones y los asuntos de los implicados o responsables (*stakeholders*) sirven como foco organizativo de la evaluación (como base para determinar qué información se necesita), que se lleva a cabo dentro de los planteamientos metodológicos del paradigma constructivista.

La utilización de las demandas, preocupaciones y asuntos de los implicados es necesaria, según Guba y Lincoln, porque:

a) Son *grupos de riesgo* ante la evaluación y sus problemas deben ser convenientemente contemplados, de manera que se sientan protegidos ante tal riesgo.

b) Los resultados pueden ser utilizados en su *contra* en diferentes sentidos, sobre todo si están al margen del proceso.

c) Son potenciales usuarios de la información resultante de la evaluación.

d) Pueden ampliar y mejorar el rango de la evaluación.

e) Se produce una interacción positiva entre los distintos implicados.

El cambio paradigmático lo justifican estos autores porque:

a) La metodología convencional no contempla la necesidad de identificar las demandas, preocupaciones y asuntos de los implicados.

b) Para llevar a cabo lo anterior se necesita una postura de descubrimiento más que de verificación, típica del positivismo.

c) No se tienen en cuenta suficientemente los factores contextuales.

d) No se proporcionan medios para valoraciones caso por caso.

e) La supuesta neutralidad de la metodología convencional es de dudosa utilidad cuando se buscan juicios de valor acerca de un objeto social.

Partiendo de estas premisas, el evaluador es responsable de determinadas tareas, que realizará secuencialmente o en paralelo, construyendo un proceso ordenado y sistemático de trabajo. Las responsabilidades básicas del evaluador de la cuarta generación son las siguientes:

1) Identificar todos los implicados con riesgo en la evaluación.

2) Resaltar para cada grupo de implicados sus construcciones acerca de lo evaluado y sus demandas y preocupaciones al respecto.

3) Proporcionar un contexto y una metodología hermenéutica para poder tener en cuenta, comprender y criticar las diferentes construcciones, demandas y preocupaciones.

4) Generar el máximo acuerdo posible acerca de dichas construcciones, demandas y preocupaciones.

5) Preparar una agenda para la negociación acerca de temas no consensuados.

6) Recoger y proporcionar la información necesaria para la negociación.

7) Formar y hacer de mediador para un «forum» de implicados para la negociación.

8) Desarrollar y elaborar informes para cada grupo de implicados sobre los distintos acuerdos y resoluciones acerca de los intere-

ses propios y de los de otros grupos (Stake, 1986; Zeller, 1987).

9) Reciclar la evaluación siempre que queden asuntos pendientes de resolución.

La propuesta de Guba y Lincoln (1989) se extiende bastante en la explicación de la naturaleza y características del paradigma constructivista en contraposición con las del positivista.

Cuando se habla de los pasos o fases de la evaluación en esta cuarta generación, sus proponentes citan doce pasos o fases, con diferentes subfases en cada una de estas. Estos pasos son los siguientes:

1) Establecimiento de un *contrato* con un patrocinador o cliente.

•Identificación del cliente o patrocinador de la evaluación.

•Identificación del objeto de la evaluación.

•Propósito de la evaluación (Guba y Lincoln, 1982).

•Acuerdo con el cliente por el tipo de evaluación.

•Identificación de audiencias.

•Breve descripción de la metodología a usar.

•Garantía de acceso a registros y documentos.

•Acuerdo por garantizar la confidencialidad y anonimato hasta donde sea posible.

•Descripción del tipo de informe a elaborar.

•Listado de especificaciones técnicas.

2) *Organización* para reciclar la investigación.

•Selección y entrenamiento del equipo evaluador.

•Consecución de facilidades y acceso a la información (Lincoln y Guba, 1985).

3) Identificación de las *audiencias* (Guba y Lincoln, 1982).

•Agentes.

•Beneficiarios.

•Victimas.

4) Desarrollo de *construcciones conjuntas* dentro de cada grupo o audiencia (Glaser y Strauss, 1967; Glaser, 1978; Lincoln y Guba, 1985).

5) *Contraste y desarrollo* de las construcciones conjuntas de las audiencias.

•Documentos y registros.

•Observación.

•Literatura profesional.

•Círculos de otras audiencias.

•Construcción ética del evaluador.

6) *Clasificación* de las demandas, preocupaciones y asuntos resueltos.

7) *Establecimiento de prioridades* en los temas no resueltos.

8) *Recogida* de información.

9) Preparación de la *agenda* para la negociación.

10) Desarrollo de la *negociación.*

11) *Informe*s (Zeller, 1987; Licoln y Guba, 1988).

12) Reciclado/*revisión*.

Para juzgar la calidad de la evaluación, se nos ofrecen tres enfoques que se denominan *paralelo,* el ligado al *proceso hermenéutico* y el de *autenticidad.*

Los criterios *paralelos,* de confianza, se denominan así porque intentan ser paralelos a los criterios de rigor utilizados muchos años dentro del paradigma convencional. Estos criterios han sido validez interna y externa, fiabilidad y objetividad. Sin embargo, los criterios deban ser acordes con el paradigma fundamentador (Morgan, 1983). En el caso de la cuarta generación los criterios que se ofrecen son los de *credibilidad, transferencia, dependencia* y *confirmación* (Lincoln y Guba, 1986).

El criterio de *credibilidad* es paralelo al de validez interna, de forma que la idea de isomorfismo entre los hallazgos y la realidad se reemplaza por el isomorfismo entre las realidades construidas de las audiencias y las reconstrucciones del evaluador atribuidas a ellas. Para conseguir esto existen varias técnicas, entre las que destacan las siguientes: a) el compromiso prolongado, b) la observación persistente, c) el contraste con colegas, d) el análisis de casos negativos (Kidder, 1981), e) la subjetividad progresiva y f) el control de los miembros. La *transferencia* puede verse como paralela a la validez externa, la *dependencia* es paralela al criterio de fiabilidad y la *confirmación* puede verse como paralela a la objetividad.

Otra manera de juzgar la calidad de la evaluación es el análisis dentro del propio proceso, algo que encaja con el *paradigma hermenéutico*, a través de un proceso dialéctico.

Pero estos dos tipos de criterios, aunque útiles, no son del todo satisfactorios para Guba y Lincoln, que defienden con más ahínco los criterios que denominan de *autenticidad,* también de base constructivista. Estos criterios incluyen los siguientes: a) imparcialidad, justicia, b) autenticidad ontológica, c) autenticidad educativa, d) autenticidad catalítica y e) autenticidad táctica (Lincoln y Guba, 1986).

Este análisis de la cuarta generación de podemos terminarlo con los rasgos con los que definen Guba y Lincoln a la evaluación:

a) La evaluación es un proceso *sociopolítico.*

b) La evaluación es un proceso conjunto de *colaboración.*

c) la evaluación es un proceso de *enseñanza/aprendizaje.*

d) La evaluación es un proceso *continuo, recursivo* y *altamente divergente.*

e) La evaluación es un proceso *emergente.*

f) La evaluación es un proceso con resultados *impredecibles.*

g) La evaluación es un proceso que *crea realidad.*

En esta evaluación, se retienen las características del evaluador fruto de las tres primeras generaciones, esto es, la de técnico, la de analista y la de juez, pero estas deben ampliarse con destrezas para recoger e interpretar datos cualitativos (Patton, 1980), con la de historiador e iluminador, con la de mediador de juicios, así como un papel más activo como evaluador en un contexto socio-político concreto.

Russell y Willinsky (1997) defienden las potencialidades del planteamiento de la cuarta generación para desarrollar formulaciones alternativas de práctica evaluadora entre los implicados, incrementando la probabilidad de que la evaluación sirva para mejorar la enseñanza escolar. Esto requiere por parte del profesorado el reconocimiento de otras posiciones, además de la suya, la implicación de todos desde el principio del proceso y, por otra parte, el desarrollo de aproximaciones más pragmáticas de la conceptualización de Guba y Lincoln, adaptadas a las distintas realidades escolares.

7. El nuevo impulso alrededor de Stufflebeam

Para terminar este recorrido analítico-histórico desde los primeros intentos de medición educativa hasta la actual investigación evaluativa en educación, queremos recoger las recomendaciones que más recientemente nos viene ofreciendo una de las figuras señeras de este campo en la segunda mitad del siglo XX.

Nos estamos refiriendo a Daniel L. Stufflebeam, proponente del modelo CIPP (el más utilizado) a finales de los sesenta, desde 1975 a 1988 presidente del «*Joint Committee on Standars for Educational Evaluation*» y actual director del «*Evaluation Center*» de la Western Michigan University (sede del Joint Committee) y del CREATE (*Center for Research on Educational Accountability and Teacher Evaluation*), centro auspiciado y financiado por el Departamento de Educación del gobierno americano.

Recogiendo estas recomendaciones (Stufflebeam, 1994, 1998, 1999, 2000 y 2001), en las que se han ido integrando ideas de diversos evaluadores también notables, no sólo ofrecemos una de las últimas aportaciones a la actual concepción de la investigación evaluativa en educación, sino que completamos en buena medida la visión del panorama actual, rico y plural, tras analizar la cuarta generación de Guba y Lincoln.

Se parte de los cuatro principios del *Joint Committee* (1981 y 1988), esto es, de la idea de que cualquier buen trabajo de investigación evaluativa debe ser: a) *útil,* esto es, proporcionar información a tiempo e influir, b) *factible,* esto es, debe suponer un esfuerzo razonable y debe ser políticamente viable, c) *apropiada, adecuada, legítima,* esto es, ética y justa con los implicados, y d) *segura* y *precisa* a la hora de ofrecer información y juicios sobre el objeto de la evaluación. Además, la evaluación se ve como una «transdisciplina», pues es aplicable a muchas disciplinas diferentes y a muchos objetos diversos (Scriven, 1994).

Stufflebeam invoca a la responsabilidad del evaluador, que debe actuar de acuerdo a principios aceptados por la sociedad y a criterios de profesionalidad, emitir juicios sobre la calidad y el valor educativo del objeto evaluado y debe asistir a los implicados en la interpretación y utilización de su información y sus juicios. Sin embargo, es también su deber, y su derecho, estar al margen de la lucha y la responsabilidad política por la toma de decisiones y por las decisiones tomadas.

Para evaluar la educación en una sociedad moderna, Stufflebeam (1994) nos dice que se deben tomar algunos criterios básicos de referencia como los siguientes:

- Las *necesidades educativas.* Es necesario preguntarse si la educación que se proporciona cubre las necesidades de los estudiantes y de sus familias en todos los terrenos a la vista de los derechos básicos, en este caso, dentro de una sociedad democrática (Nowakowski y otros, 1985).

- La *equidad.* Hay que preguntarse si el sistema es justo y equitativo a la hora de proporcionar servicios educativos, el acceso a los mismos, la consecución de metas, el desarrollo de aspiraciones y la cobertura para todos los sectores de la comunidad (Kellagan, 1982).

- La *factibilidad.* Hay que cuestionar la eficiencia en la utilización y distribución de recursos, la adecuación y viabilidad de las normas legales, el compromiso y participación de los implicados y todo lo que hace que el esfuerzo educativo produzca el máximo de frutos posibles.

- La *excelencia* como objetivo permanente de búsqueda. La mejora de la calidad, a partir del análisis de las prácticas pasadas y presentes es uno de los fundamentos de la investigación evaluativa.

Tomando el referente de estos criterios y sus derivaciones, Stufflebeam sumariza una serie de recomendaciones para llevar a cabo buenas investigaciones evaluativas y mejorar el sistema educativo. Estas recomendaciones son las siguientes:

1) Los planes de evaluación deben satisfacer los cuatro requerimientos de *utilidad, factibilidad, legitimidad* y *precisión* (Joint Committee, 1981 y 1988).

2) Las entidades educativas deben examinarse por su integración y servicio a los *principios de la sociedad democrática,* equidad, bienestar, etc.

3) Las entidades educativas deben ser valoradas tanto por su *mérito* (valor intrínseco, calidad respecto a criterios generales) como por su *valor* (valor extrínseco, calidad y servicio para un contexto particular) (Guba y Lincoln, 1982; Scriven, 1991), como por su *significación* en la realidad del contexto en el que se ubica. Scriven (1998) nos señala que usando otras denominaciones habituales, mérito tiene bastante equivalencia con el término calidad, valor con el de relación coste-eficacia y significación con el de importancia. En todo caso, los tres conceptos son dependientes del contexto, sobre todo significación, de manera que entender la diferencia entre dependencia del contexto y arbitrariedad es parte de la comprensión de la lógica de la evaluación.

4) La evaluación de profesores, instituciones educativas, programas, etc, debe relacionarse siempre con el conjunto de sus deberes, responsabilidades y obligaciones profesionales o institucionales, etc. Quizás uno de los retos que deben abordar los sistemas educativos es la definición más clara y precisa de estos *deberes y responsabilidades*. Sin ello, la evaluación es problemática, incluso en el terreno formativo (Scriven, 1991*a*).

5) Los estudios evaluativos deben ser capaces de valorar hasta qué medida los profesores y las instituciones educativas son *responsables* y *rinden cuentas* del cumplimiento de sus deberes y obligaciones profesionales (Scriven, 1994).

6) Los estudios evaluativos deben proporcionar *direcciones para la mejora*, porque no basta con emitir un juicio sobre el mérito o el valor de algo.

7) Recogiendo los puntos anteriores, todo estudio evaluativo debe tener un componente *formativo* y otro *sumativo.*

8) Se debe promover la *autoevaluación* profesional, proporcionando a los educadores las destrezas para ello y favoreciendo actitudes positivas hacia ella (Madaus y otros, 1991)

9) *La evaluación del contexto* (necesidades, oportunidades, problemas en un área,...) debe emplearse de manera *prospectiva,* para localizar bien las metas y objetivos y definir prioridades. Asimismo, la evaluación del contexto debe utilizarse *retrospectivamente,* para juzgar bien el valor de los servicios y resultados educativos, en relación con las necesidades de los estudiantes (Madaus y otros, 1991; Scriven, 1991)

10) *La evaluación de las entradas* (inputs) debe emplearse de manera *prospectiva,* para asegurar el uso de un rango adecuado de enfoques según las necesidades y los planes.

11) *La evaluación del proceso* debe usarse de manera *prospectiva* para mejorar el plan de trabajo, pero también de manera *retrospectiva* para juzgar hasta qué punto la calidad del proceso determina el por qué los resultados son de un nivel u otro (Stufflebean y Shinkfield, 1987).

12) *La evaluación del producto* es el medio para identificar los resultados buscados y no buscados en los participantes o afectados por el objeto evaluado. Se necesita una valoración *prospectiva* de los resultados para orientar el proceso y detectar zonas de necesidades. Se necesita una evaluación *retrospectiva* del producto para poder juzgar en conjunto el mérito y el valor del objeto evaluado (Scriven, 1991; Webster y Edwards, 1993; Webster y otros, 1994).

13) Los estudios evaluativos se deben apoyar en la *comunicación* y en la *inclusión* sustantiva y funcional de los implicados (*stakeholders*) con las cuestiones claves, criterios, hallazgos e implicaciones de la evaluación, así como en la promoción de la aceptación y el uso de sus resultados (Chelimsky, 1998). Más aún, los estudios evaluativos deben conceptualizarse y utilizarse sistemáticamente como parte del proceso de mejora educativa a largo plazo (Alkin y otros, 1979; Joint Committee, 1988; Stronge y Helm, 1991; Keefe, 1994) y de fundamento para la acción contra las discriminaciones sociales (Mertens, 1999).

(Mertens, 1999). La *evaluación para el desarrollo* (*empowerment evaluation*), que defiende Fetterman (1994), es un procedimiento, de base democrática, de participación de los implicados en el programa evaluado, para promover la autonomía de los mismos en la resolución de sus problemas. Weiss (1998) nos alerta de que la evaluación participativa incrementa la probabilidad de que se utilicen los resultados de la evaluación, pero también la de que sea conservadora en su concepción, pues es difícil pensar que los responsables de una organización pongan en cuestión el fundamento y el sistema de poder de la misma. Generalmente su interés es el cambio de cosas pequeñas.

14) Los estudios evaluativos deben emplear *múltiples perspectivas, múltiples medidas de resultados,* y métodos tanto *cuantitativos* como *cualitativos* para recoger y analizar la información. La pluralidad y complejidad del fenómeno educativo hace necesario emplear enfoques múltiples y multidimensionales en los estudios evaluativos (Scriven, 1991)

15) Los estudios evaluativos deben ser evaluados, incluyendo *metaevaluaciones formativas* para mejorar su calidad y su uso y *metaevaluaciones sumativas* para ayudar a los usuarios en la interpretación de sus hallazgos y proporcionar sugerencias para mejorar futuras evaluaciones (*Joint Committee*, 1981 y 1988; Madaus y otros, 1991; Scriven, 1991; Stufflebeam, 2001).

Estas quince recomendaciones proporcionan elementos esenciales para un enfoque de los estudios evaluativos que Stufflebeam denomina *objetivista* y que se basa en la teoría ética de que la bondad moral es objetiva e independiente de los sentimientos personales o meramente humanos.

Sin entrar en el debate sobre estas valoraciones finales de Stufflebeam, ni en análisis comparativos con otras propuestas, por ejemplo con las de Guba y Lincoln (1989), nos resulta evidente que las concepciones de la investigación evaluativa son diversas, dependiendo del origen epistemológico desde el que se parte, pero apareciendo claros y contundentes algunos elementos comunes a todas las perspectivas como la *contextualización*, el *servicio a la sociedad*, la *diversidad metodológica*, la *atención, respeto y participación de los implicados*, etc., así como una mayor *profesionalización* de los evaluadores y una mayor *institucionalización* de los estudios (Worthen y Sanders, 1991).

El propio Stufflebeam (1998) reconoce el conflicto de los planteamientos del *Joint Committee on Standards for Educational Evaluation* con las posiciones de la corriente evaluadora denominada postmodernista, representada, además de por Guba y Lincoln, por otros reconocidos evaluadores como Mabry, Stake y Walker, pero no acepta que existan razones para actitudes de escepticismo y frustración con las prácticas evaluadoras actuales, porque existen muchos ámbitos de aproximación y el desarrollo de estándares de evaluación es perfectamente compatible con la atención a los diversos implicados, valores, contextos sociales y métodos. Stufflebeam defiende una mayor colaboración en la mejora de las evaluaciones, estableciendo los estándares de manera participativa, pues cree que es posible la aproximación de planteamientos, con contribuciones importantes desde todos los puntos de vista.

Weiss (1998) también toma posiciones parecidas cuando nos dice que las ideas constructivistas deben hacernos pensar más cuidadosamente al usar los resultados de las evaluaciones, sintetizarlas y establecer generalizaciones, pero duda que todo haya que interpretarlo en términos exclusivamente individuales, pues existen muchos elementos comunes entre las personas, los programas y las instituciones.

8. Para concluir: síntesis de enfoques modélicos y metodológicos de la evaluación y la última perspectiva de Scriven

Tras este análisis del desarrollo de la evaluación a lo largo del Siglo XX, parece oportuno, a modo de síntesis y de conclusión, recoger y resaltar los que son considerados los principa-

les modelos, planteamientos metodológicos, diseños, perspectivas y visiones de la evaluación en la actualidad. Su análisis, de manera compacta, es un complemento necesario para una visión como la histórica que, por su linealidad, tiene el riesgo de ofrecer una imagen disciplinar artificiosamente fraccionada.

Hemos visto que en la década de los setenta y en sus alrededores se produce una especie de eclosión de propuestas evaluativas, que tradicionalmente han venido siendo denominadas como *modelos* (Castillo y Gento, 1995) y en algunos casos como *diseños* (Arnal y otros, 1992) de investigación evaluativa. Sabemos que existieron varias decenas de estas propuestas, pero muy concentradas en el tiempo, en la década citada. De hecho, el asunto de los propuestos modelos para la evaluación parece un tema prácticamente cerrado desde hace cuatro lustros. Ya no surgen nuevos modelos o propuestas, salvo alguna excepción como vemos más adelante.

A pesar de lo dicho, se sigue hablando de modelos, métodos y diseños en la literatura especializada, sobre todo buscando su clasificación de acuerdo con diversos criterios, origen paradigmático, propósito, metodología, etc. También en las clasificaciones, no sólo en los modelos, existe diversidad, lo que prueba que, además de dinamismo académico en el terreno de la investigación evaluativa, todavía existe cierta debilidad teórica al respecto.

Nosotros ya hemos señalado con anterioridad (Escudero, 1993), que coincidimos con Nevo (1983 y 1989) en la apreciación de que muchos de los acercamientos a la conceptualización de la evaluación (por ejemplo, el modelo respondiente, el libre de metas, el de discrepancias, etc.) se les ha denominado indebidamente como modelos a pesar de que ninguno de ellos tenga el grado de complejidad y de globalidad que debería acarrear el citado concepto. Lo que un texto clásico en evaluación (Worthen y Sanders, 1973) denomina como «modelos contemporáneos de evaluación» (a los conocidos planteamientos de Tyler, Scriven, Stake, Provus, Stufflebeam, etc), el propio Stake (1981) dice que sería mejor denominarlo como «persuasiones» mientras que House (1983) se refiere a «metáforas».

Norris (1993) apunta que el concepto de modelo se utiliza con cierta ligereza al referirse a concepción, enfoque o incluso método de evaluación. De Miguel (1989), por su parte, piensa que muchos de los llamados modelos solamente son descripciones de procesos o aproximaciones a programas de evaluación. Darling-Hammond y otros (1989) utilizan el término «modelo» por costumbre, pero indican que no lo hacen en el sentido preciso que tiene el término en las ciencias sociales, esto es, apoyándose en una estructura de supuestos interrelacionales fundamentada teóricamente. Finalmente diremos que el propio autor del modelo CIPP, solamente utiliza esta denominación de manera sistemática para referirse a su propio modelo (Stufflebeam y Shinkfield, 1987), utilizando los términos de enfoque, método, etc., al referirse a los otros. Para nosotros, quizás sea el término *enfoque* evaluativo el más apropiado, aunque aceptemos seguir hablando de modelos y diseños por simple tradición académica.

Nuestra idea es que a la hora de plantearnos una investigación evaluativa, no contamos todavía con unos pocos modelos bien fundamentados, definidos, estructurados y completos, entre los que elegir uno de ellos, pero sí tenemos distintos enfoques modélicos y un amplio soporte teórico y empírico, que permiten al evaluador ir respondiendo de manera bastante adecuada a las distintas cuestiones que le va planteando el proceso de investigación, ayudándole a configurar un *plan global,* un *organigrama coherente,* un «modelo» científicamente robusto para llevar a cabo su evaluación (Escudero, 1993). ¿Cuáles son las cuestiones que hay que responder en este proceso de construcción modélica? Apoyándonos en las aportaciones de diferentes autores (Worthen y Sanders, 1973; Nevo, 1989; Kogan, 1989; Smith y Haver, 1990), deben responderse y delimitar su respuesta al construir un modelo de investigación evaluativa, los aspectos siguientes:

1) Objeto de la investigación evaluativa.
2) Propósito, objetivos.
3) Audiencias/implicados/clientela.
4) Énfasis/aspectos prioritarios o preferentes.
5) Criterios de mérito o valor.
6) Información a recoger.
7) Métodos de recogida de información.
8) Métodos de análisis.
9) Agentes del proceso.
10) Secuenciación del proceso.
11) Informes/utilización de resultados.
12) Límites de la evaluación.
13) Evaluación de la propia investigación evaluativa / metaevaluación.

Para definir estos elementos hay que buscar, lógicamente, el apoyo de los diferentes enfoques modélicos, métodos, procedimientos, etc., que la investigación evaluativa ha desarrollado, sobre todo en las últimas décadas.

Volviendo a los denominados modelos de los setenta y a sus clasificaciones, podemos recoger algunas de las aparecidas en la última década en nuestro entorno académico, apoyándose en distintos autores. Así, por ejemplo, Arnal y otros (1992) ofrecen una clasificación de lo que denominan *diseños de la investigación evaluativa,* revisando las de diversos autores (Patton, 1980; Guba y Lincoln, 1982; Pérez, 1983; Stufflebeam y Shinkfield, 1987). Esta clasificación es la siguiente :

Tabla 1- Tipos de diseños de investigación educativa

Perspectiva	*Patton (1980)*	*Guba y Lincoln (1982)*	*Pérez (1983)*	*Stufflebeam y Shinkfield (1987)*	*Autores creadores*
Empírico-analítica	Objetivos	Objetivos	Objetivos	Objetivos	Tyler (1950)
	Análisis sistemas		Análisis sistemas		Rivlin (1971) Rossi y otros (1979)
				Método científico	Suchman (1967)
Susceptibles de complementariedad	CIPP	CIPP	CIPP		Stufflebeam (1966)
	Crítica artística	Crítica artística	Crítica artística		Eisner (1971)
	Adversario			Contrapuesto	Wolf (1974)
	UTOS	UTOS	Cronbach (1982)		
Humanístico-interpretativa	Respondente	Respondente	Respondente	Respondente	Stake (1975
	Iluminativo		Iluminativo	Iluminativo	Parlett y Hamilton (1977)
	Sin metas	Sin metas		Sin metas	Scriven (1967)
			Democrático		MacDonald (1976)

Por su parte, Castillo y Gento (1995) ofrecen una clasificación de «métodos de evaluación» dentro de cada uno de los modelos (paradigmas), que ellos denominan conductivista-eficientistas, humanísticos y holísticos (mixtos). Una síntesis de estas clasificaciones es la siguiente:

Tabla 2- Modelo conductista-eficientista

Método/ autor	*Finalidad evaluativa*	*Paradigma dominante*	*Contenido de evaluación*	*Rol del evaluador*
Consecución objetivos Tyler (1940)	Medición logro objetivos	Cuantitativo	Resultados	Técnico externo
CIPP Stufflebeam (1967)	Información para toma decisiones	Mixto	C (contexto) I (input) P (proceso) P (producto)	Técnico externo
Figura (countenance) Stake (1967)	Valoración resultados y proceso	Mixto	Antecedentes, transacciones, resultados	Técnico externo
CSE Alkin (1969)	Información para determinación de decisiones	Mixto	Centrados en logros de necesidades	Técnico externo
Planificación educativa Cronbach (1982)	Valoración proceso y producto	Mixto	U (unidades de evaluación) T (tratamiento) O (operaciones)	Técnico externo

Tabla 3- Modelo humanístico

Método/ autor	*Finalidad evaluativa*	*Paradigma dominante*	*Contenido de evaluación*	*Rol del evaluador*
Atención al cliente Scriven (1973)	Análisis de necesidades del cliente	Mixto	Todos los efectos del programa	Evaluador externo de necesidades del cliente
Contraposición Owens (1973), Wolf (1974)	Opiniones para decisión consensuada	Mixto	Cualquier aspecto del programa	Árbitro externo del debate
Crítica artística Eisner (1981)	Interpretación crítica de la acción educativa	Cualitativo	• Contexto • Procesos emergentes • Relaciones de procesos • Impacto en contexto	Provocador externo de interpretaciones

Tabla 4- Modelo holístico

Método/ autor	*Finalidad evaluativa*	*Paradigma dominante*	*Contenido de evaluación*	*Rol del evaluador*
Evaluación respondente Stake (1976)	Valoración de respuesta a necesidades de participantes	Cualitativo	Resultado de debate total sobre programa	Promotor externo de la interpretación por los implicados
Evaluación holística MacDonald (1976)	Interpretación educativa para mejorarla	Cualitativo	Elementos que configuran la acción educativa	Promotor externo de la interpretación por los implicados
Evaluación iluminativa Parlett y Hamilton (1977)	Iluminación y comprensión de los componentes del programa	Cualitativo	Sistema de enseñanza y medio de aprendizaje	Promotor externo de la interpretación por los implicados

También Scriven (1994) ofrece una clasificación de los «modelos anteriores», previamente a introducir su perspectiva transdisciplinar que luego comentamos. Este autor identifica *seis visiones* o enfoques alternativos en la fase «explosiva» de los modelos, además de algunas más que denomina «exóticas» y que se mueven entre los modelos de jurisprudencia y de experto. A continuación comentamos sucintamente estas visiones y los «modelos» que se adscriben a ellas.

La visión fuerte hacia la toma de decisiones (Visión A) concibe al evaluador investigando con el objetivo de llegar a conclusiones evaluativas que le ayuden al que debe tomar decisiones. Los que apoyan este enfoque se preocupan de si el programa alcanza sus objetivos, pero van más allá, cuestionándose si tales objetivos cubren las necesidades que deben cubrir. Esta posición es mantenida, aunque no la hiciera explícita, por Ralph Tyler y extensamente elaborada en el modelo CIPP (Stufflebeam y otros, 1971).

Según el planteamiento tyleriano, las decisiones acerca de un programa deben basarse en el grado de coincidencia entre los objetivos y los resultados. El cambio de los alumnos, habitualmente el objetivo perseguido, es el criterio de evaluación.

A diferencia de Tyler, Stufflebeam ofrece una perspectiva más amplia de los contenidos a evaluar. Estos son las cuatro dimensiones que identifican su modelo, *contexto* (C) donde tiene lugar el programa o está la institución, *inputs* (I) elementos y recursos de partida, *proceso* (P) que hay que seguir hacia la meta y *producto* (P) que se obtiene. Además, se deja constancia de que el objetivo primordial de la investigación evaluativa es la mejora, la toma de decisiones para la mejora de todas y cada una de las cuatro dimensiones antes citadas.

Scriven (1994) nos dice que Stufflebeam ha seguido desarrollando su perspectiva desde que desarrolló el CIPP. Sin embargo, uno de sus colaboradores en tal empresa, Guba, tomó posteriormente una dirección diferente, tal como hemos visto al analizar la cuarta generación de la evaluación (Guba y Lincoln, 1989).

La visión débil hacia la toma de decisiones (Visión B) concibe al evaluador proporcionando información relevante para la toma de decisiones, pero no le obliga a emitir conclusiones evaluativas o críticas a los objetivos de los programas. El representante teórico más genui-

no es Marv Alkin (1969), que define a la evaluación como un proceso factual de recogida y generación de información al servicio del que toma las decisiones, pero es éste el que tiene que tomar las conclusiones evaluativas. Esta posición es lógicamente popular entre los que piensan que la verdadera ciencia no debe o no puede entrar en cuestiones de juicios de valor. El modelo de Alkin se conoce como CES (Centro para el Estudio de la Evaluación), planteando las siguientes fases: valoración de las necesidades y fijación del problema, planificación del programa, evaluación de la instrumentalización, evaluación de progresos y evaluación de resultados.

La *visión relativista* (Visión C) también mantiene la distancia de las conclusiones evaluativas, pero usando el marco de valores de los clientes, sin un juicio por parte del evaluador acerca de esos valores o alguna referencia a otros. Esta visión y la anterior han sido el camino que ha permitido integrarse sin problemas en el «carro» de la investigación evaluativa a muchos científicos sociales. De hecho, uno de los textos más utilizados de evaluación en el ámbito de las ciencias sociales (Rossi y Freeman, 1993), toma preferentemente esta perspectiva .

Las visiones B y C son las posiciones de los científicos entroncados con una concepción libre de valores de la ciencia. En cambio, los que participan de la visión A proceden de un paradigma diferente, probablemente debido a su conexión académica con la historia, la filosofía de la educación, la educación comparada y la administración educativa.

Hace unos años Alkin (1991) revisó sus planteamientos de dos décadas atrás, pero siguió sin incluir los términos de mérito, valor o valía; termina definiendo un Sistema de Información para la Gestión (*Management Information System*-MIS) para uso del que toma decisiones, pero no ofrece valoraciones al respecto

Pero la forma más simple de la *visión relativista* (Visión C) es la desarrollada en el «modelo de discrepancia» de evaluación de Malcolm Provus (1971). Las discrepancias son las divergencias con la secuencia de tareas proyectadas y la temporalización prevista. Este modelo es muy cercano al control de programas en sentido convencional; es una especie de simulación de una evaluación.

La visión de la descripción fértil, rica, completa (Visión D) es la que entiende la evaluación como una tarea etnográfica o periodística, en la que el evaluador informa de lo que ve sin intentar emitir afirmaciones valorativas o inferir conclusiones evaluativas, ni siquiera en el marco de los valores del cliente como en la visión relativista. Esta visión ha sido defendida por Robert Stake y muchos de los teóricos británicos. Se trata de una especie de versión naturalista de la visión B, tiene algo de sabor relativista y a veces parece precursora de la visión de la cuarta generación. Se centra en la observación, en lo observable, más que en la inferencia. Recientemente se le ha denominado como *visión de la descripción sólida, fuerte,* para evitar el término *rica,* que parece más evaluativa.

Stake, en su primera etapa, es tayleriano en cuanto a concepción evaluativa centrada en los objetivos planteados, proponiendo el método de evaluación de la *figura* (Stake, 1967), como rostro o imagen total de la evaluación. Esta gira en torno a los tres componentes, *antecedentes, transacciones* y *resultados,* elaborando dos matrices de datos, una de *descripción* y otra de *juicio*. En la primera se recogen de un lado las *intenciones* y de otro las *observaciones* y, en la segunda, las *normas,* lo que se aprueba y los *juicios,* lo que se cree que debe ser.

A mitad de los setenta, Stake se aleja de la tradición tayleriana de preocupación por los objetivos y revisa su método de evaluación hacia un planteamiento que él califica como *«respondente»* (Stake, 1975 y 1975*a*), asumiendo que los objetivos del programa pueden modificarse sobre la marcha, con la finalidad de ofrecer una visión completa y holística del programa y *responder* a los problemas y cuestiones reales que plantean los implicados. Según Stufflebeam y Shinkfield (1987), este modelo hizo de Stake el líder de una nueva escue-

la de evaluación, que exige un método pluralista, flexible, interactivo, holístico, subjetivo y orientado al servicio. Este modelo sugiere la «atención al cliente» propuesta por Scriven (1973), valorando sus necesidades y expectativas.

De manera gráfica, Stake (1975*a*) propone las fases del método a modo de las horas de un reloj, poniendo la primera en las doce horas y siguiendo las siguientes fases, el sentido de las agujas del reloj. Estas fases son las siguientes: 1) Hablar con los clientes, responsables y audiencias, 2) Alcance del programa, 3) Panorama de actividades, 4) Propósitos e intereses, 5) Cuestiones y problemas, 6) Datos para investigar los problemas, 7) Observadores, jueces e instrumentos, 8) Antecedentes, transacciones y resultados, 9) Desarrollo de temas, descripciones y estudio de casos, 10) Validación (confirmación), 11) Esquema para la audiencia y 12) Reunión de informes formales. El evaluador puede seguir las fases también en sentido contrario del reloj o en cualquier otro orden.

En el método respondente el evaluador ha de entrevistar a los implicados para conocer sus puntos de vista y buscar la confluencia de las diversas perspectivas. El evaluador deberá interpretar las opiniones y diferencias de puntos de vista (Stecher y Davis, 1990) y presentar una amplia gama de opiniones o juicios, en lugar de presentar sus conclusiones personales.

La visión del proceso social (Visión E) que cristalizó hace algo más de dos décadas alrededor de un grupo de la Universidad de Stanford, dirigido por Lee J. Cronbach (1980), resta importancia a la orientación sumativa de la evaluación (decisiones externas sobre los programas y rendición de cuentas), enfatizando la *comprensión,* la *planificación* y la *mejora* de los programas sociales a los que sirve. Sus posiciones quedaban claramente establecidas en noventa y cinco tesis que han tenido una enorme difusión entre los evaluadores y los usuarios de la evaluación.

En cuanto a los contenidos de la evaluación, Cronbach (1983) propone que se planifiquen y controlen los siguientes elementos:

- *Unidades* (U) que son sometidas a evaluación, individuos o grupos participantes.
- *Tratamiento* (T) de la evaluación.
- *Operaciones* (O) que lleva a cabo el evaluador para la recogida y análisis de datos, así como para la elaboración de conclusiones.
- *Contexto* en el que tiene lugar el programa y su evaluación.

En una investigación evaluativa concreta se pueden dar varias unidades, varios tratamientos y varias operaciones, en definitiva varios (uto), dentro de un universo UTO de situaciones admisibles.

Ernie House (1989), un teórico y un práctico de la evaluación bastante independiente de corrientes en boga, también marcó el entronque social de los programas, pero se distinguía sobre todo por su énfasis de las dimensiones más éticas y argumentales de la evaluación, quizás motivado por la ausencia de estas facetas en los planteamientos de Cronbach y sus colaboradores.

La *visión constructivista de la cuarta generación* (Visión F) es la última de estas seis visiones que describe Scriven (1994), siendo mantenida por Guba y Lincoln (1989) y seguida por muchos evaluadores americanos y británicos. Ya hemos visto anteriormente que esta visión rechaza una evaluación orientada a la búsqueda de calidad, mérito, valor, etc., y favorece la idea de que ello es el resultado de la construcción por individuos y la negociación de grupos. Esto significa, según Scriven, que el conocimiento científico de todo tipo es sospechoso, discutible y no objetivo. Lo mismo le ocurre a todo trabajo análitico como el análisis filosófico, incluido el suyo. Scriven apunta que el propio Guba ha sido siempre consciente de las potenciales «autocontradicciones» de su posición.

De esta revisión de Scriven quedan al margen algunas posiciones evaluativas tradicionalmente recogidas y tratadas por los analistas. Así por

ejemplo, Schuman (1967) ofrece un diseño evaluativo basado en el *método científico* o, al menos, en alguna variación o adaptación del mismo. Owens (1973) y Wolf (1974 y 1975) proponen un método de *contraposición* o discusión que sobre un programa llevan a cabo dos grupos de evaluadores, partidarios y adversarios, para proporcionar información pertinente a quienes toman decisiones. Eisner (1971, 1975 y 1981) plantea la evaluación en términos similares al proceso de crítica artística.

El propio Scriven (1967 y 1973) proponía hace años centrar la evaluación en la *atención al cliente* y no tanto en las metas previstas, puesto que frecuentemente los logros no previstos son más importantes que los que figuran en la planificación del programa. Por ello, se suele denominar a su enfoque como *evaluación sin metas*. El evaluador determina el valor o mérito del programa para informar a los usuarios; se trata algo así como de un intermediario informativo (Scriven, 1980).

La evaluación iluminativa (Parlett y Hamilton, 1977) tiene un enfoque holístico, descriptivo e interpretativo, con la pretensión de iluminar sobre un complejo rango de cuestiones que se dan de manera interactiva (Fernández, 1991). *La evaluación democrática* de MacDonald (1971 y 1976), también denominada holística, supone la participación colaborativa de los implicados, siendo el contraste de opiniones de los implicados el elemento evaluativo primordial.

Scriven (1994) analiza las seis visiones críticamente y se muestra más cercano a la visión A, *la visión fuerte sobre la toma de decisiones*, representada fundamentalmente por el modelo CIPP de Stufflebeam y sus planteamientos, pues dice que es la más cercana de todas a la *visión del sentido común*, que es la que tienen los evaluadores trabajando con sus programas, de la misma manera que los médicos trabajan con los pacientes, haciéndolo lo mejor posible, independientemente del tipo y del estado general del paciente. Scriven quiere extender esta visión con una visión o modelo que denomina *transdisciplinar* y que él califica como significativamente distinta de la aludida visión A y radicalmente diferente de las restantes.

En la *perspectiva transdisciplinar,* la investigación evaluativa tiene dos componentes: el conjunto de campos de aplicación de la evaluación y el contenido de la propia disciplina. Algo parecido a lo que ocurre a disciplinas como la estadística y la medición. En definitiva, la investigación evaluativa es una disciplina que incluye sus propios contenidos y los de otras muchas disciplinas; su preocupación por el análisis y mejora se extiende a muchas disciplinas, es transdisciplinar.

Esta visión es *objetivista* como la A y defiende que el evaluador determine el mérito o el valor del programa, del personal o de los productos investigados. En tal sentido, se debe establecer de manera explícita y defender la lógica utilizada en la inferencia de conclusiones evaluativas a partir de las premisas definicionales y factuales. Así mismo, se deben perseguir las falacias argumentales de la doctrina libre de valores (Evaluation Thesaurus, 1991).

En segundo lugar, la *perspectia transdisciplinar* se orienta hacia el consumidor, más que hacia el gestor o intermediario. No se trata de una orientación exclusiva hacia el consumidor, pero sí la consideración primera del consumidor como justificación del programa, y que el bien común es la primacía de la evaluación. A partir de aquí, también se produce información valiosa para el gestor que decide y se pueden analizar los productos de un programa o institución a la vista de sus objetivos. Esta posición no sólo ve legitimidad en la emisión de conclusiones evaluativas por parte del investigador, sino que ve necesidad de hacerlo en la gran mayoría de las ocasiones.

Se trata también de una *visión generalizada,* no justamente una visión general, que incluye la generalización de conceptos en el ámbito del conocimiento y la práctica. Desde esta perspectiva, la investigación evaluativa es mucho más que la evaluación de programas e incide en procesos, instituciones y otros muchos más

objetos. De manera más detallada, esta visión generalizada significa que:

a) Los campos distintivos de aplicación de la disciplina son los programas, el personal, los rendimientos, los productos, los proyectos, la gestión, y la metaevaluación de todo ello.

b) Las investigaciones evaluativas inciden en todo tipo de disciplinas y en las prácticas que resultan de ellas.

c) Las investigaciones evaluativas se mueven desde niveles muy prácticos hasta el nivel conceptual.

d) Los distintos campos de la investigación evaluativa tienen muchos niveles de interconexión y solapamiento. La evaluación de programas, de personal, de centros, etc., tienen muchos puntos en común.

El cuarto elemento distintivo de la visión transdisciplinar de la evaluación es que se trata de una visión *técnica*. La evaluación no sólo necesita el apoyo técnico de otras muchas disciplinas, sino que, además, tiene su propia metodología. La lógica de la síntesis de resultados, las consecuencias, etc., y la correcta ubicación en el proceso de muchas técnicas auxiliares en las que, probablemente, no es necesario ser un gran especialista, pero sí tener un conocimiento cabal.

Esta *perspectiva transdisciplinar* de la investigación evaluativa de Scriven (1994), coincide en gran medida con los planteamientos que de la misma hemos defendido en otros momentos (Escudero, 1996). Nosotros no tenemos unas posiciones contrarias a las otras visiones en la misma medida que las tiene Scriven y, de hecho, consideramos desde una *posición pragmática,* que todas las visiones tienen puntos fuertes y que en todo caso, aportan algo útil para la comprensión conceptual y el desarrollo de la investigación evaluativa. Sin embargo, sí que pensamos que esta moderna visión de Scriven es sólida y coherente y ampliamente aceptada en la actualidad.

Una crítica que podría hacerse a este planteamiento de Scriven está en el excesivo énfasis relativo de la orientación al cliente, al usuario en sentido estricto. Pensamos que esta orientación debe integrarse dentro de una orientación a los implicados, donde existen distintos tipos y distintas audiencias y, por supuesto, una muy importante, son los usuarios en el sentido de Scriven, pero nos parece que la investigación evaluativa hoy en día tiene una *orientación prioritaria* más plural que la defendida por este autor.

Referencias bibliográficas

Ahman, S. J. y Cook, M. D. (1967). Evaluating Pupil Growth. Principles of Tests Measurement. Boston, Ma.: Allyn and Bacon

Alkin, M. (1969). Evaluation theory development. Evaluation Comment, 2, 1, 2-7.

Alkin, M. (1991). Evaluation theory development: II. En M. McLaughlin, y Phillips (Eds.), Evaluation and education at quarter century (pp.91-112). Chicago: NSSE/ University of Chicago .

Anderson, S. B. y Ball, S. (1983). The profession and practice of program evaluation. San Francisco, Ca: Jossey-Bass.

Arnal, J.; Del Rincon, D. y Latorre, A. (1992). Investigación educativa. Fundamentos y Metodología. Barcelona: Labor

Atkin, J. M. (1968). Behavioral Objectives in curriculum design: A cautionary note. The Science Teacher, 35, 27-30.

Baker, L. R. (1969). Curriculum evaluation. Review of Educational Research, 39, 339-358.

Barbier, J. M. (1993). La evaluación en los procesos de formación. Barcelona: Paidos.

Berk, A. R. (Ed.) (1981). Educational evaluation methodology: The state of the art. Baltimore: The Hopkins University Press.

Blanco, F. (1994). La evaluación en la educación secundaria. Salamanca: Amasú Ediciones.

Bloom, B. S., Engelhaut, M.D., Furst, E.J., Hill, W.H. y Krathwohl, D.R. (1956). Taxonomy of educational objectives Handbook I; Cognitive domain. New York: Davis, McKay.

Bloom, B. S., Hastings, J. Th. y Madaus, F. (1971). Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill.

Bloom, B.S., Hastings, T. y Madaus G. (1975). Evaluación del aprendizaje. Buenos Aires: Troquel.

Bonboir, A. (1972). La Docimologie. Paris: PUF.

Cabrera, F. (1986). Proyecto docente sobre técnicas de medición y evaluación educativas. Barcelona: Universidad de Barcelona.

Carreño, H. F. (1977). Enfoques y principios teóricos de la evaluación. Mexico: Trillas.

Castillo, S. y Gento, S. (1995). Modelos de evaluación de programas educativos. En A.Medina y L. M. Willar (Coord.), Evaluación de programas educativos, centros y profesores (pp.25-69). Madrid: Editorial Universitas, S. A..

Chelimsky, E. (1998). The role of experience in formulating theories of evaluation practice. American Journal of Evaluation 19, 1, 35-55.

Coffman, W. E. (1971). Essay examinations. En R.L. Thorndike (Ed.) Educational Measurement. Washington , DC: American Council on Education.

Cronbach, L. J. (1963). Course improvement through evaluation. Teachers College Record, 64, 672-683.

Cronbach, L. J. (1982). Designing evaluations of educational and social programs. Chicago: Jossey-Bass.

Cronbach, L. J., Hambron, S.R., Dornbusch, S.M., Hess, R.D., Hornick, R.C., Phillips, D.C., Walker, D.F. y Weiner, S.S. (1980). Towards reform in program evaluation: Aims, methods and institutional arrangements. San Francisco: Jossey-Bass.

Cronbach, L. J. y Suppes, P. (1969). Research for tomorrow's schools: Disciplined inquiry for education. New York: MacMillan.

Darling-Hammond, L., Wise, AE, & Pease, SR (1989). Teacher evaluation in the organizational context: A review of the literature. En E. R. House (Ed.) New directions in educational evaluation (pp.203-253). London: The Falmer Press.

De la Orden, A. (1985). Investigación evaluativo. En Arturo De la Orden. (Ed.), Investigación educativa. Diccionario de Ciencias de la Educación (pp.133-137). Madrid: Anaya.

De Miguel, M. (1989). Modelos de investigación sobre organizaciones educativas. Revista de Investigación Educativa, 7, 13, 21-56.

Dubois, P. H. (1970). A History of Psychological Testing. Boston: Allyn Bacon.

Ebel, R. L. (1977). Fundamentos de la medición educacional. Buenos Aires: Guadalupe.

Eisner, E. W. (1967). Educational objectives: Help or hindrance?. The School Review, 75, 250-260.

Eisner, E. W. (1969, Instructional and expressive educational objectives: their formulation and use in curriculum. En J. Popham (Ed.), Instructional objectives (pp. 1-18). Chicago: AERA.

Eisner, E. (1971). Emerging models for educational evaluation. School Review, 2.

Eisner, E. W. (1975). The perceptive eye: Toward the reformation of educational evaluation. Stanford, Ca: Stanford Evaluation Consortion.

Eisner, E. W. (1981). The methodology of qualitative evaluation: the case of educational connoisseurship and educational criticism. Stanford, Ca: Stanford University .

Eisner, W. E. (1985). The art of educational evaluation. London: The Falmer Press.

Escudero, T. (1993). Enfoques modélicos en la evaluación de la enseñanza universitaria, Actas de las III Jornadas Nacionales de Didáctica Universitaria «Evaluación y Desarrollo Profesional» (pp. 5-59). Las Palmas: Servicio de Publicaciones, Universidad de Las Palmas.

Escudero, T. (1996). Proyecto docente e investigador. Zaragoza: Universidad de Zaragoza.

Fernández Ballesteros, R. (1981). Perspectivas históricas de la evaluación conductual. En R. Fernández, y J.A.I Carrobles. (Ed.), Evaluación conductual. Madrid: Ediciones Pirámide.

Fernández de Castro, J. (1973). La enseñanza programada. Madrid: CSIC.

Fernández, J. (1991). La evaluación de la calidad docente. En A. Medina (Coord.), Teoría y métodos de evaluación. Madrid: Cincel.

Fetterman, D. M. (1994). Empowerment evaluation. Evaluation Practice, 15, 1, 1-15.

Gagne, R. M. (1971). Las condiciones del aprendizaje. Madrid: Aguilar.

Gil, E. (1992). El sistema educativo de la Compañía de Jesús. La «Ratio Studiorum». Madrid: UPCO.

Glaser, B. G. (1978). Theoretical sensitivity. Mill Valley, Ca: Sociology Press.

Glaser, B. G. y Strauss, A. L. (1967). The discovery of grounded theory. Chicago: Aldine.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: some questions. American Psychologists, 18, 519-521.

Glaser, R. (Dir.) (1965). Teaching machines and programmed learning. Washington: National Education Association.

Gronlund, N. E. (1985). Measurement and evaluation in teaching. New York: MacMillan,

Guba, G. E. y Lincoln, Y. S. (1982). Effective evaluation. San Francisco: Jossey Bass Publishers.

Guba, E. G. y Lincoln, Y. S. (1989). Fourth Generation Evaluation. Newbury Park, Ca.: Sage Publications

Gullicksen, H. (1950). Theory of mental tests. New York: Wiley

Hambleton, R. K. (1985). Criterion-referenced measurement.En Encyclopedia of Educational Research. New York: McMillan.

Hammond, R. L. (1983). Evaluation at the local level. En B. R. Worthen y J. R. Sanders, Educational Evaluation: Theory and Practice. Worthington, Ohio: Charles A. Jones Publishing Company.

Hernández, F. (1993). Proyecto docente e investigador. Murcia.

Horowitz, R. (1995). A 75-year legacy on assessment: Reflections from an interview with Ralph W. Tyler. The Journal of Educational Research, 89, 2, 68-75.

House, E. R. (1983). How we think about evaluation. En House, E. R., Philosophy of evaluation. San Francisco, Ca.: Jossey-Bass

House, E. R. (1989). Evaluating with validity. Newbury Park, Ca.: Sage.

Joint Committee on Standards for Educational Evaluation (1981, Standards for evaluations of educational programs, projects, and materials. New York.: McGraw-Hill.

Joint Committee on Standards for Educational Evaluation (1988). The personnel evaluation standards. Newbury Park, CA.: Sage.

Keefe, J. (1994). School evaluation using the CASE-IMS model and improvement process. Studies in Educational Evaluation, 20,1, 55-67.

Kellaghan, T. (1982). La evaluación educativa. Bogotá: Universidad Pontificia Javierana.

Kidder, L. H. (1981). Qualitative research and quasi-experimental frameworks. En M. B. Brewer y B. E. Collins, (Eds.), Scientific inquiry and the social sciences. San Francisco: Jossey-Bass.

Kogan, M. (Ed.) (1989). Evaluating higher education. London: Jessica Kingsley Publishers.

Krathwohl, R. D., Blomm, B.S. y Masia, B.B. (1964). Taxonomy of educational objectives. Handbook II: Affective domain. New Cork: Davis McKay.

Lewy, A. (Ed.) (1976). Manual de evaluación formativa del currículo. Bogotá : Voluntad/Unesco.

Lincoln, y. S. y Guba, E. G. (1985). Naturalistic inquiry. Beverly Hills, CA: Sage.

Lincoln, y. S. y Guba, E. G. (1986). But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. en D. D. Williams (Ed.), Naturalistic evaluation. San Francisco: Jossey-Bass.

Lincoln, y. S. y Guba, E. G. (1988). Criteria for assessing naturalistic inquiries as products. Paper presented at the American Educational Research Association, New Orleans, LA.

Lindquist, E. F. (1953). Design and Analysis of Experiments in Psychology and Education. Boston, Ma.: Houghton Mifflin Company.

Lindvall, C. M., (Ed.) (1964). Defining educational objectives. Pittsburg: University of Pittsburgh Press.

Macdonald, B. (1971). The evaluation of the Humanities Curriculum Project: a holistic approach. Theory into Practice, 10, 3, 163-169.

Macdonald, B. (1976). Evaluation and the control of education. En D. Tawney (Ed.), Curriculum evaluation today: trends and implications, 125-136. London: McMillan.

Madaus, G. F. y otros (1991). Evaluation Models. Viewpoints on Educational and Human Services Evaluation. Hingham , Mass.: Kluwer-Nijhoff Publishing.

Mager, R. F. (1962). Preparing instructional objective. Palo Alto, CA.: Fearon.

Mager, R. F. (1973). Análisis de metas. México: Trillas.

Mann, H., (1845). Boston Grammar and Writing Schools. Common School Journal, October, 7, 19.

Martínez de Toda, M. J. (1991). Metaevaluación de necesidades educativas: Hacia un sistema de normas. Tesis doctoral. Madrid: Universidad Complutense.

Mateo, J. (1986). Proyecto docente e investigador. Barcelona: Universidad de Barcelona.

Mateo, J. y otros (1993). La evaluación en el aula universitaria. Zaragoza: ICE-Universidad de Zaragoza.

Mccall, W. A. (1920). A new kind of school examination. Journal of Educational Research, January.

McReynold, P. (1975). Advances in Psychological Assessment, vol. III.San Francisco: Jossey-Bass.

Mertens, D. M. (1999). Inclusive evaluation: Implications of transformative theory for evaluation. American Journal of Evaluation, 20, 1, 1-14.

Metfessel, N. S. y Michael, W. B. (1967). A paradigm involving multiple criterion measures for the evaluation of effectiveness of school programs. Educational and Psychological Measurement, 27, 931-943.

Morgan, G. (1983). Beyond method. Beverly Hills, Ca: Sage.

Nevo, D. (1983). The conceptualization of educational evaluation: An analytical review of the literature. Review of Educational Research, 53, 1, 117-128.

Nevo, D. (1989). The conceptualization of educational evaluation: An analytical review of the literature. En E. R. House (Ed.), New directions in educational evaluation. London : The Falmer Press, 15-29.

Norris, N. (1993). Understanding educational evaluation. London : Kogan Page/CARE, School of Education , University of East Anglia .

Nowalowski, Jeri, Mary Anne Bunda, Russell Working (1985). A Handbook of Educational Variables. Boston: Kluwer-Nijhoff .

Nunnally, J. C. (1978). Psychometric Theory. New York: McGraw Hill.

Owens, T. R. (1973). Educational evaluation by adversery proceedings. En E.R. House (Comp.), School Evaluation: The Politics and Process. Berkeley: McCutchan.

Parlett, M. y Hamilton , d. (1977). Evaluation as illumination: A new approach to the study of innovative programmes. En Hamilton D. y otros (Eds.), Beyond the numbers game. London: MacMillan.

Patton, M. Q. (1980). Qualitative Evaluation methods. Beverly Hills, Ca.: Sage

Pérez, A. (1983). Modelos contemporáneos de evaluación. En J. Gimeno y A. Pérez, La enseñanza: su teoría y su práctica (pp. 426-449). Madrid: Akal

Phillips, R. C. (1974). Evaluation in education. Columbus, Ohio: Merril.

Pieron, H. (1968). Vocabulaire de la psychologie. Paris: PUF.

Pieron, H. (1969). Examens et Docimologie. PUF, Paris.

Planchard, E. (1960). La investigación pedagógica. Madrid: Ediciones Fas.

Popham, W. J. (1970). Establishing instructional goals. Englewood Cliffs, N. J.: Prentice Hall.

Popham, W. J. (1980). Problemas y técnicas de la evaluación educativa. Madrid: Anaya.

Popham, W. J. (1983). Evaluación basada en criterios. Madrid: Magisterio Español, S. A..

Popham, W. y Baker, E. L. (1970). Systematic Instruction. Englewood Cliffs, NJ.: Prentice Hall.

Provus, M. (1971). Discrepancy evaluation. For educational program improvement and assessment. Berkeley, Ca.: McCutchan Publishing Co.

Rivlin, A. M. (1971). Systematic thinking for social action. Washington, D.C.: Brookings Institute.

Rodríguez, T. y otros (1995). Evaluación de los aprendizajes. Aula Abierta Monografías 25. Oviedo: ICE-Universidad de Oviedo.

Rosenthal, J. E. (1976). Evaluation history. En S. B. Anderson, y otros (Eds.), Encyclopedia of Educational Evaluation. San Francisco: Jossey Bass Publishers.

Rossi, P. H. y Freeman, H. (1993). Evaluation: A Systematic Approach. Beverly Hills , Ca.: Sage.

Rossi, P. H. y otros (1979). Evaluation: A systematic approach. Beverly Hills , Ca.: Sage.

Russell, N. y Willinsky, J. (1997). Fourth generation educational evaluation: The impact of a post-modern paradigm on school based evaluation. Studies in Educational Evaluation, 23, 3, 187-199.

Rutman, L. (Ed.) (1984). Evaluation research methods: A base guide. Beverly Hills, Ca.: Sage.

Rutman, L. y Mowbray, G. (1983). Understanding program evaluation. Beverly Hills, Ca.: Sage.

Salvador, L. (1992). Proyecto docente., Universidad de Cantabria.

Scriven, M. (1967). The methodology of evaluation. En Perspectives of Curriculum Evaluation, (pp. 39-83). AERA Monograph 1. Chicago : Rand McNally and Company.

Scriven, M. (1973). Goal-free evaluation. En E. R. House (Ed.), School evaluation: The politics and process,(pp. 319-328). Berkeley, CA.: McCutchan.

Scriven, M. (1974). Prose and cons about goal-free evaluation. Evaluation Comment, 3, 1-4.

Scriven, M. (1980). The logic of evaluation. Inverness , Ca.: Edgepress

Scriven, M. (1991, Evaluation Thesaurus. Newbury Park , Ca.: Sage

Scriven, M. (1991a). Duties of the teacher. Kalamazoo, Mi.: Center for Research on Educational Accountability and Teacher Evaluation.

Scriven, M. (1994). Evaluation as a discipline. Studies in Educational Evaluation, 20, 1, 147-166.

Scriven, M. (1998). Minimalist theory: The least theory that practice require. American Journal of Evaluation 19, 1, 57-70.

Smith, E. R. y Tyler, R. W. (1942). Appraising and recording student progress. New York: Harper & Row,.

Smith, N. L. y Haver, d. M. (1990). The applicability of selected evaluation models to evolving investigative designs. Studies in Educational Evaluation, 16, 3, 489-500.

Stacher, B. M. y Davis, W. A. (1990). How to Focus on Evaluation. Newbury Park, Ca.: Sage.

Stake, R. E. (1967). The countenance of educational evaluation. Teacher College Record, 68, 523-540.

Stake, R. E. (1975a). Program evaluation: particularly responsive evaluation. Occasional Paper, 5. University of Western Michigan.

Stake, R. E. (1975b). Evaluating the arts in education: A responsive approach. Ohio: Merril .

Stake, R. E. (1976). A theoretical stament of responsive evaluation. Studies in Educational Evaluation, 2 , 19-22.

Stake, R. E. (1981). Setting standars for educational evaluators. Evaluation News, 22, 148-152.

Stake, R. E. (1986). Quieting reform. Urbana: University of Illinois Press.

Stenhouse, L. (1984). Investigación y desarrollo del curriculum. Madrid: Morata.

Stronge, J. H. y Helm, V. M. (1991). Evaluating professional support personnel in educational settings. Newbury Park, Ca: Sage.

Stufflebeam, D. L. (1966). A depth study of the evaluation requeriment. Theory into Practice, 5, 3, 121-134.

Stufflebeam, D. L. (1994). Introduction: Recommendations for improving evaluations in U. S. public schools. Studies in Educational Evaluation, 20, 1, 3-21.

Stufflebeam, D. L. (1998). Conflicts between standards-based and postmodernist evaluations: Toward rapprochement. Journal of Personnel Evaluation in Education, 12, 3, 287-296.

Stufflebeam, D. L. (1999). Using profesional standards to legally and ethically release evaluation findings. Studies in Educational Evaluation, 25, 4, 325-334.

Stufflebeam, D. L. (2000). Guidelines for developing evaluation checklists. Consultado en www.wmich.edu/evalctr/checklists/ el 15 de Diciembre de 2002.

Stufflebeam, D. L. (2001). The metaevaluation imperative. American Journal of Evaluation, 22, 2, 183-209.

Stufflebeam, D. L., Foley, WJ, Gephart, WJ, Guba, EG, Hammond, RL, Merriman, HO & Provus, MM (1971). Educational Evaluation and Decision-making, Itasca, Illinois : F. E. Peacock Publishing.

Stufflebeam, D. L. y Shinkfield, A. J. (1987). Evaluación sistemática. Guía teórica y práctica. Barcelona: Paidos/MEC.

Suchman, E. A. (1967). Evaluative Research: Principles and Practice in Public Service and Social Action Programs. New York : Russell Sage Foundation.

Sunberg, N. D. (1977). Assessment of person. Englewood Cliffs, N.J.: Prentice Hall.

Taba, H. (1962). Curriculum development. Theory and practice. New York: Harcourt Brace .

Thorndike, E. L,. (1904). An Introduction to the Theory of Mental and Social Measurements. New York: Teacher College Press, Columbia University .

Tyler, R. W. (1950). Basic principles of curriculum and instruction. Chicago: University of Chicago Press .

Tyler, R. W. (1967). Changing concepts of educational evaluation. En R. E. Stack (Comp.), Perspectives of curriculum evaluation. AERA Monograph Series Curriculum Evaluation, 1. Chicago: Rand McNally,.

Tyler, R. W., (Ed.) (1969). Educational evaluation: New roles, new means. Chicago: University of Chicago Press .

Walberg, H. J. y Haertel, G. D. (Ed.) (1990). The International Encyclopedia of Educational Evaluation. Oxford: Pergamon Press.

Webster, W. J. y Edwards, M. E. (1993). An accountability system for school improvement, Paper presented at the annual meeting (April) of the AERA. Atlanta, GA.

Webster, W. J., Mendro, R.L. y Almaguer, T.O. (1994). Effectiveness indices: A «value added» approach to measuring school effect. Studies in Educational Evaluation, 20, 1, 113-137.

Weiss, C. H. (1983). Investigación evaluativa. Métodos para determinar la eficiencia de los programas de acción. México: Trillas,.

Weiss, C. H. (1998). Have we learned anything new about the use of evaluation?. American Journal of Evaluation, 19, 1, 21-33.

Wilson, A. R. J. (1978). La evaluación de los objetivos. En J. A. R. Wilson (Ed.), Fundamentos psicológicos del aprendizaje y la enseñanza (pp. 549-578). Madrid: Anaya..

Wolf, R. L. (1974). The citizen as jurist: A new model of educational evaluation. Citizen Action in Education, 4.

Wolf, R. L. (1975). Trial by jury: a new evaluation method. Phi Delta Kappa, 57, 185-187.

Worthen, B. R. y Sanders, J. R. (1973). Educational Evaluation: Theory and Practice. Worthington, Ohio: Charles a. Jones Publishing Company.

Worthen, B. R. y Sanders, J. R. (1991). The changing face of educational evaluation, Theory into Practice, XXX, 1, 3-12.

Zeller, N. C. (1987). A rethoric for naturalistic inquiry. Ph. D. dissertation, Indiana University.

ABOUT THE AUTHORS / SOBRE LOS AUTORES

Tomás Escudero Escorza (tescuder@posta.unizar.es). Catedrático del área de Métodos de Investigación y Diagnóstico en Educación de la Universidad de Zaragoza. Es miembro de la Comisión de Coordinación Técnica del Plan de Calidad de las Universidades. Tiene un gran número de trabajos y publicaciones sobre evaluación educativa, particularmente sobre evaluación institucional.

ARTICLE RECORD / FICHA DEL ARTÍCULO

Reference / Referencia	Escudero, Tomás (2003). Desde los tests hasta la investigación evaluativa actual. Un siglo, el XX, de intenso desarrollo de la evaluación en educación. *Revista ELectrónica de Investigación y EValuación Educativa* (RELIEVE), *v. 9*, n. 1. http://www.uv.es/RELIEVE/v9n1/RELIEVEv9n1_1.htm. Consultado en (*poner fecha*).
Title / Título	Desde los tests hasta la investigación evaluativa actual. Un siglo, el XX, de intenso desarrollo de la evaluación en educación. [*From tests to current evaluative research. One century, the XXth, of intense development of evaluation in education*]
Authors / Autores	Tomás Escudero Escorza
Review / Revista	Revista ELectrónica de Investigación y EValuación Educativa (RELIEVE), *v. 9*, n. 1
ISSN	1134-4032
Publication date / Fecha de publicación	2003 (**Reception Date**: 2003 January 10; **Publication Date**: 2003 Febr. 27)
Abstract / Resumen	*This article presents a review and state of art about development in educational evaluation in the XXth century. The main theoretical proposals are commented.* Este artículo presenta una revisión crítica del desarrollo histórico que ha tenido el ámbito de la evaluación educativa durante todo el siglo XX. Se analizan los principales propuestas teóricas planteadas..
Keywords Descriptores	*Evaluation, Evaluation Research, Evaluation Methods; Formative Evaluation Summative Evaluation, Testing, Program Evaluation.* Evaluación, Investigación evaluativa, Métodos de Evaluación, Evaluación Formativa, Evaluación Sumativa, Test, Evaluación de Programas
Institution / Institución	Universidad de Zaragoza (España)
Publication site / Dirección	http://www.uv.es/RELIEVE
Language / Idioma	Español (Title, abstract and keywords in english)

Revista ELectrónica de Investigación y EValuación Educativa (RELIEVE)

[ISSN: 1134-4032]

© Copyright 2002, RELIEVE. Reproduction and distribution of this articles it is authorized if the content is no modified and their origin is indicated (RELIEVE Journal, volume, number and electronic address of the document). / © Copyright 2002, RELIEVE. Se autoriza la reproducción y distribución de este artículo siempre que no se modifique el contenido y se indique su origen (RELIEVE, volumen, número y dirección electrónica del documento).